PERCEPTION OF SPATIAL INFORMATION IN A MULTIPLE OBJECT AUDITORY SPACE

By

Do Manh Anh, B.Sc., B.E. (Hons)

1977

A Thesis

presented for the Degree of Doctor of Philosophy

in Electrical Engineering at the University of Canterbury,

Christchurch, New Zealand.

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisor, Professor L. Kay, Head of the Electrical Engineering Department, University of Canterbury, for his guidance and encouragement during the time I have worked on this project.

I acknowledge the help of Mr W.K. Kennedy and Mr F.M. Cady, senior lecturers in Electrical Engineering, who assisted me in the early stages of the project in using the EAI 590 Hybrid Computer for system simulation.

I would also like to thank Messrs C.H. Rowe, A. Vernon and H.J. Cusdin, senior technicians of the Electrical Engineering Department, who gave me precious technical advice in building experimental equipment; and all the subjects who participated in the psychophysical experiments.

I do appreciate all the helpful discussions with staff and postgraduate students of the Electrical Engineering Department in this research.

The scholarship granted by the Colombo Plan for my study is gratefully acknowledged.

Finally, I am grateful to my father, sisters and brother for their encouragement and support during my research.

ABSTRACT

Auditory space produced by means of a Continuous Transmission Frequency Modulated (CTFM) wide beam wide band sonar has been used in a sensory aid for the blind for some time. Underwater sonar using this auditory space has also shown its increasing promise in fish finding applications. Yet, laboratory measurements on the perception of spatial information produced by this auditory display were conducted under stationary conditions only, and showed a poor resolution capability of subjects which is entirely unmatched with the observed performance of either the sensory aid user or the fishing sonar operator in his mobility or shoal tracking task. In this thesis, laboratory experiments using realistic system simulation are described which determine the ability of subjects to detect and resolve objects in space by auditory means under both static and dynamic conditions. Step by step investigation on the auditory perception of spatial information is approached from the very simple situation of two single stationary tones to the extremely complicated situation of two multiple component tones varying with respect to time in a noise contaminated auditory space. It is shown that the auditory resolution and detectability are significantly improved in the dynamic case, when the inputs to the subjects' ears are very rich in information.
In addition to these experiments, theoretical study on the auditory display, the sonar system and several related factors are also extensively developed in this thesis.

PREFACE

The thesis covers several topics related to either psychophysics or electrical engineering system concept, or both subjects at the same time. Psychophysical experiments using system simulation techniques are described in chapters 2, 4, 5, 6 and 8, while chapters 3 and 7 discuss the audible information produced by the wide band binaural sonar and compare it with the results obtained from field experiments. Many parts of chapters 2, 3 and 4 were published in the two papers:

[1] KAY, L. and DO, M.A. An artificially generated multiple object auditory space for use where vision is impaired. Acustica, Vol.36, 1976/77, pp.1-8.

[2] DO, M.A. and KAY, L. Resolution in an artificially generated multiple object auditory space using new auditory sensations. Acustica, Vol.36, 1976/77, pp.9-15.

CONTENTS

Page

APPENDICES (Cont'd)

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 PROJECT HISTORY

It was first proposed by Kay [1] that an auditory space for use by blind persons could be artificially generated by employing a wide angle ultrasonic radiating field as an "illuminating" source, together with two receivers, one feeding each ear, to convert reflections into audible sounds. These would be perceived binaurally. The distance to a reflection was to be coded in the form of rising pitch with increasing distance (frequency proportional to distance), and the binaural differences, which could include time, frequency and amplitude, would indicate the direction of a reflection. Subsequently, Kay [2] found that amplitude difference was the dominant one. It was later demonstrated by Rowell [3] that time and frequency differences were incompatible, impeding the fusion of binaural signals, and he proposed that the possible interaural differences be reduced to that of amplitude only - as far as this was physically possible. This auditory space was readily produced by means of a sensory aid for the blind [4,5] (Fig. 1), and a sonar to locate fish [6,7] (Fig. 2).

An extensive evaluation of the sensory aid was carried out during 1970-72. The results obtained from field experiments with both systems are briefly described in section 1.2.

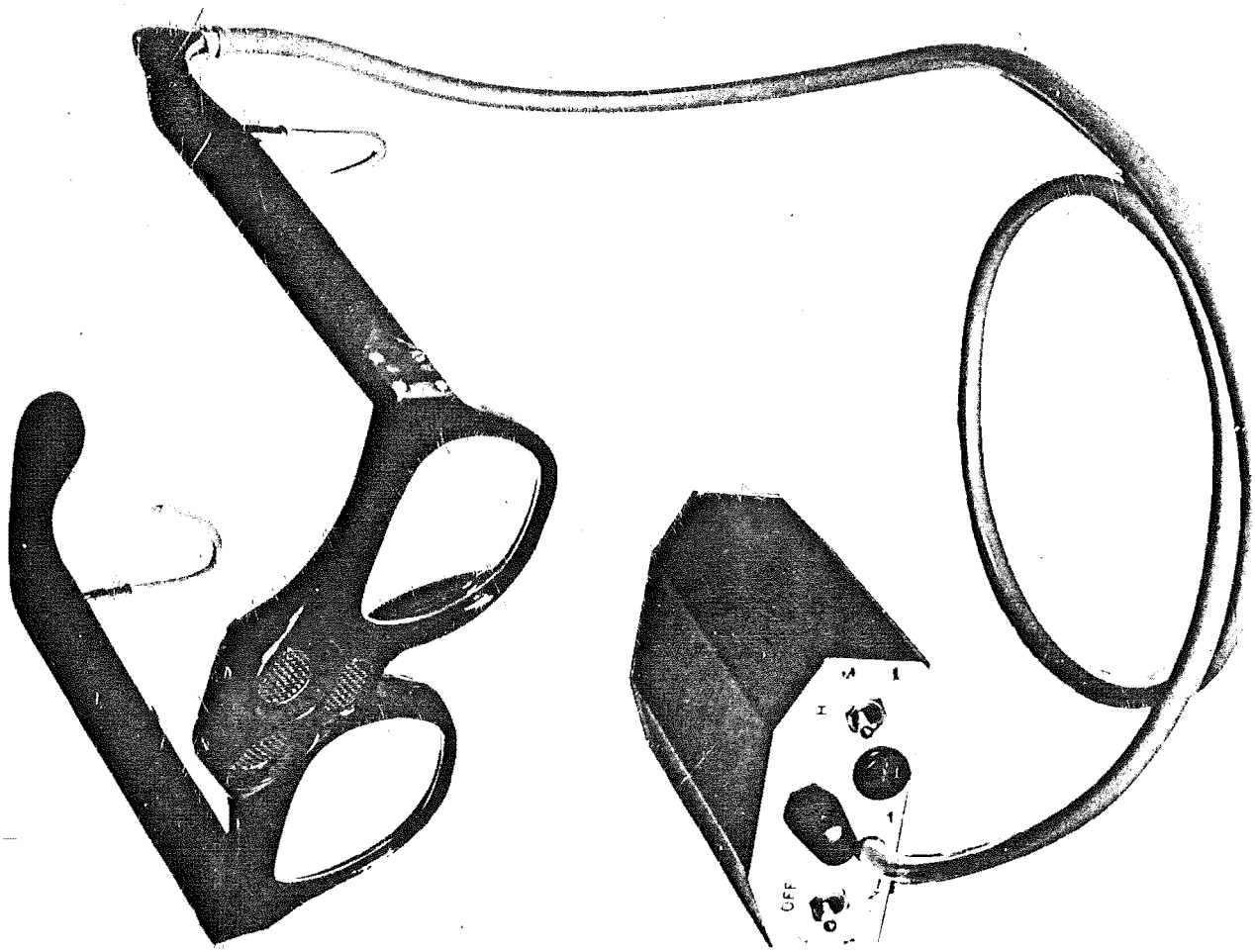


Fig. 1: Binaural sensory aid for the blind.

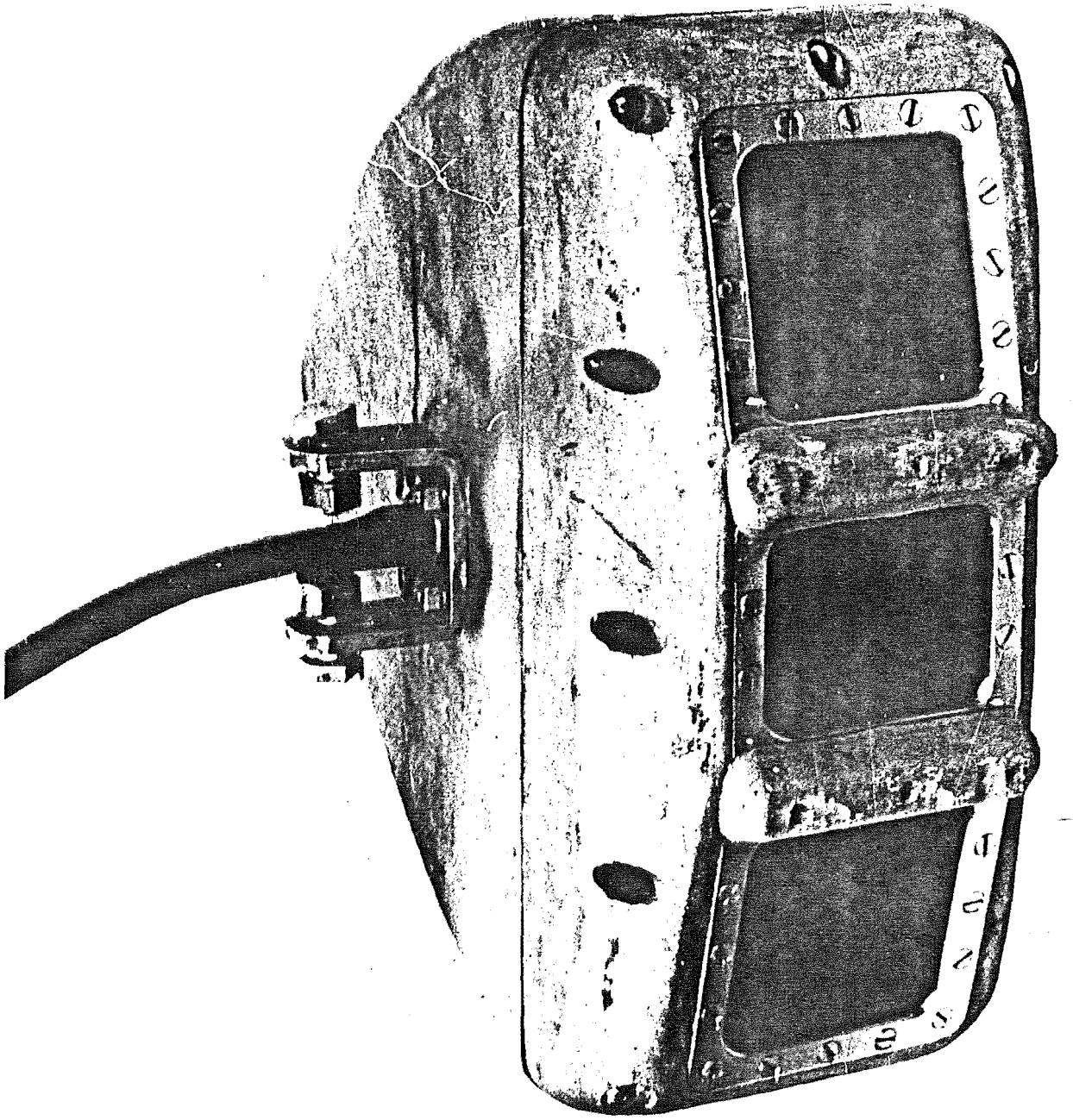


Fig. 2: Transducer for binaural sonar.

The observed ability of a blind user of the sensory aid to negotiate complex, real environments and be more effectively mobile [8,9], as well as the ability of a sonar operator to discriminate between two shoals of fish simultaneously presented in the field of view [7], were impossibly explained by the experimental results obtained from Rowell [3] and Anke [10]. The latter was from laboratory experiments using system simulation designed to determine the ability of subjects to resolve objects in space by auditory means. Rowell and Anke found a very poor auditory resolution.

It is the disagreement between the results obtained from the field evaluation [7,8,9] and laboratory experiments [3,10] which has triggered this project. Rowell and Anke's experiments used only static situations to facilitate controlled measurement. The temporal effects and the complexness of the sounds produced by either the sensory aid or the fishing sonar under dynamic conditions, which could possibly enhance the resolution capability of subjects, were neglected in their experiments.

Hence the aim of this project is to study the perception of spatial information of subjects using the multiple auditory space under dynamic conditions. Laboratory experiments using system simulation as close to realistic situations as possible have been carried out to measure the capability of resolution, detection, etc. of subjects. This result is connnected with the theoretical research on the auditory system and the sonar system to explain and predict the performance of the man and machine system as a whole.

1.2 EVALUATION OF THE BINAURAL SONAR SYSTEM

1.2.1 Sensory Aid for the Blind

About 200 blind people have learned to use the sensory system as an aid to spatial perception and to gain improved mobility. When 100 people had been trained by specially qualified teachers, both the users and teachers were questioned by a professionally prepared questionnaire [8,9]. It was found that the users learned to interpret the very rich auditory information with comparative ease. After 40 to 60 hours of training over a period of four weeks, many of the users were able to travel in a busy pedestrian area with a grace and confidence not seen before in totally blind people. A few, at least, approached a behaviour pattern not unlike that of a sighted pedestrian (as recorded on film).

Several schools specialising in the training of mobility instructors, two of which were involved in the evaluation,* have now commenced training more instructors in the use of the binaural sensory aid. This form of auditory space has now become accepted as a viable means for aiding blind people to perceive their immediate environment.

1.2.2 Fishing Sonar

The fishing sonar was fitted to the bow of a 5.4 metre boat and tested in 36 to 54 metres of water in the Marlborough Sounds (N.Z.). Even when the water was choppy with 0.3 m waves causing the boat to rock and pitch considerably, it was easy to track a 45 cm diameter sphere submerged to 18 m as the boat and sphere drifted apart up to 100 m.

---

*Department of Blind Rehabilitation, College of Education, Western Michigan University, Kalamazoo, Michigan 49001, U.S.A.; and National Guide Dog Training Centre, Royal Guide Dogs for the Blind Associations of Australia, P.O. Box 162, Kew, Victoria 3101, Australia.

The transducer was not stabilised; its field of view covers an arc of 50° in azimuth and 30° in elevation. The listener was able to correlate the movement of the boat with the changes in auditory sounds and "stabilise" his auditory space. This was the most unexpected part of the findings using the new display. Kay's earlier trials with a stabilised FM sonar had demonstrated the capability of the auditory system to track similar underwater targets [11].

In later sea trials lasting four days on a 14 m launch, again with the transducer fitted to the bows, shoals of fish could be readily detected up to 135 m. The boat was easily steered by an untrained person over the top of the shoal so as to obtain confirmatory evidence on an echo sounder. When two shoals were located at the same time, in different directions and at different distances, the choice could be made to steer over either the larger or the nearer shoal. It was also found that two different species of fish could be recognized as "different" by the difference in the character of the echo sounds they produced. The probability of locating a shoal appeared to be increased significantly [7].

Recent extensive trials (August 1975 - February 1977) at different waters of New Zealand showed great improvements on detection ranges. Shoals of fish at ranges up to 600 m were detected without difficulty (see chapter 7).

The results of both the evaluations indicated clearly the ability of human operators to use the new auditory display with relative ease, even though the input to the ears was rich in information. Whilst the sonar operator steering a boat travelling at 2.5 m/s may have up to 4 minutes in which to carry out his operation, a blind person walking at about 5 km/hour along a footpath

has only 2 secs in which to avoid a stationary pedestrian in his path at a 3 m distance. The blind traveller then experiences exceptionally high rates of fractional range change. Even so, the control functions are readily generated by the user so that the task is easily executed [12].

1.3 SYSTEM CONCEPT

The basic parameters of the auditory display of spatial information are simply described under stationary conditions. The distance to an object in the system's field of view, of say 60°, is proportional to the frequency of the audible sound output of the device (typically 941 Hz/m for the blind aid; and selectively varying among 53.33, 26.67, 13.33 and 6.67 Hz/m for the fishing sonar), and the direction of an object is indicated by the binaural difference of the sounds fed to each ear (typical I.A.D. is from 0.4 to 0.5 dB per degree in the sensory aid, and about 1.5 dB per degree in the fishing sonar). Figure 3a,b,c illustrates the basic parameters for sonar. The I.A.D. curve in Figure 3b belongs to the sensory aid. That of the fish sonar is about three times steeper [7].

To provide these combined distance and direction indicators, a continuous transmission frequency modulated (CTFM) echo-location system must be used, typically - but not restricted to - operating over a frequency range of approximately 40 kHz to 80 kHz as shown in Figure 4. There seems to be no other simple way to obtain these display parameters which are quite unique in their more complex form under dynamic conditions of human location.

The radiated signal of the transmitter is cyclic and of the form (Figure 5)

$$S_T(t) = A \exp 2\pi j \left(f_2 t_s - \frac{m}{2} t_s^2\right) \qquad (1)$$

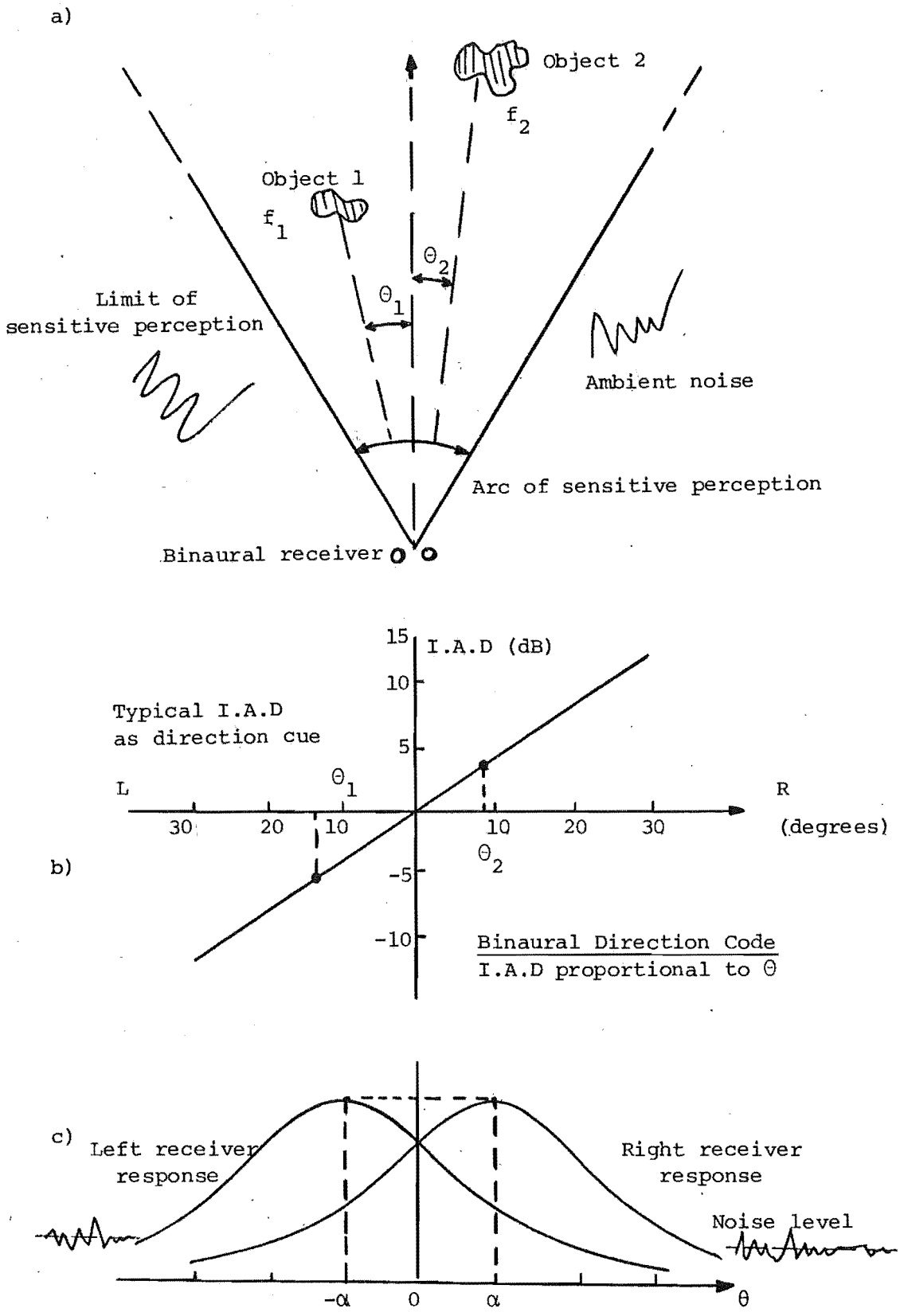


Fig. 3: Illustration of basic parameter for sonar. Maximum range is 6 m for blind aid and up to 750 m for fish sonar, the maximum audio frequency is 5 kHz - a) targets in sonic beam, b) binaural direction code, c) the receiver responses.

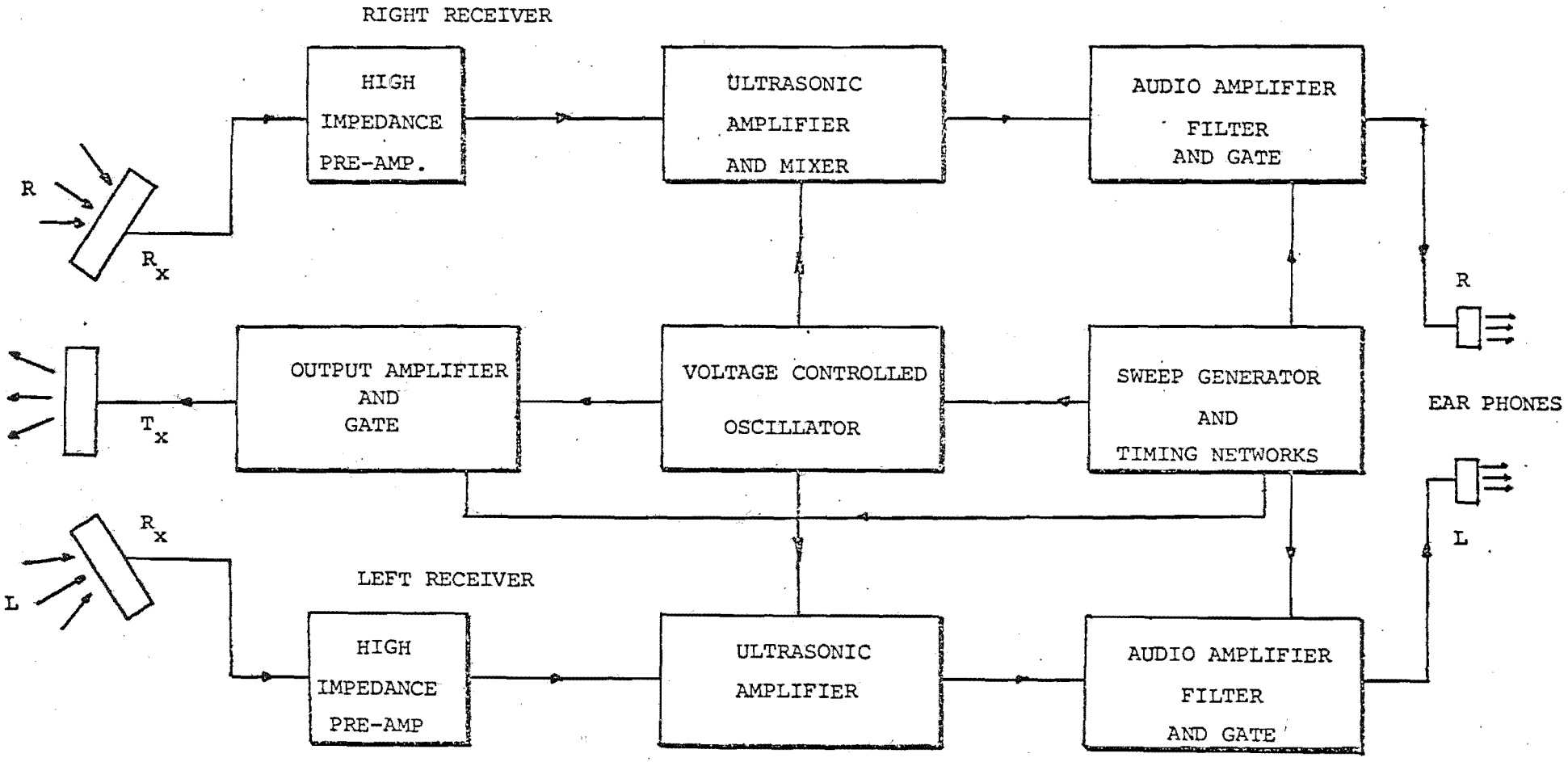


FIG.4. BLOCK DIAGRAM OF SONAR SYSTEM.

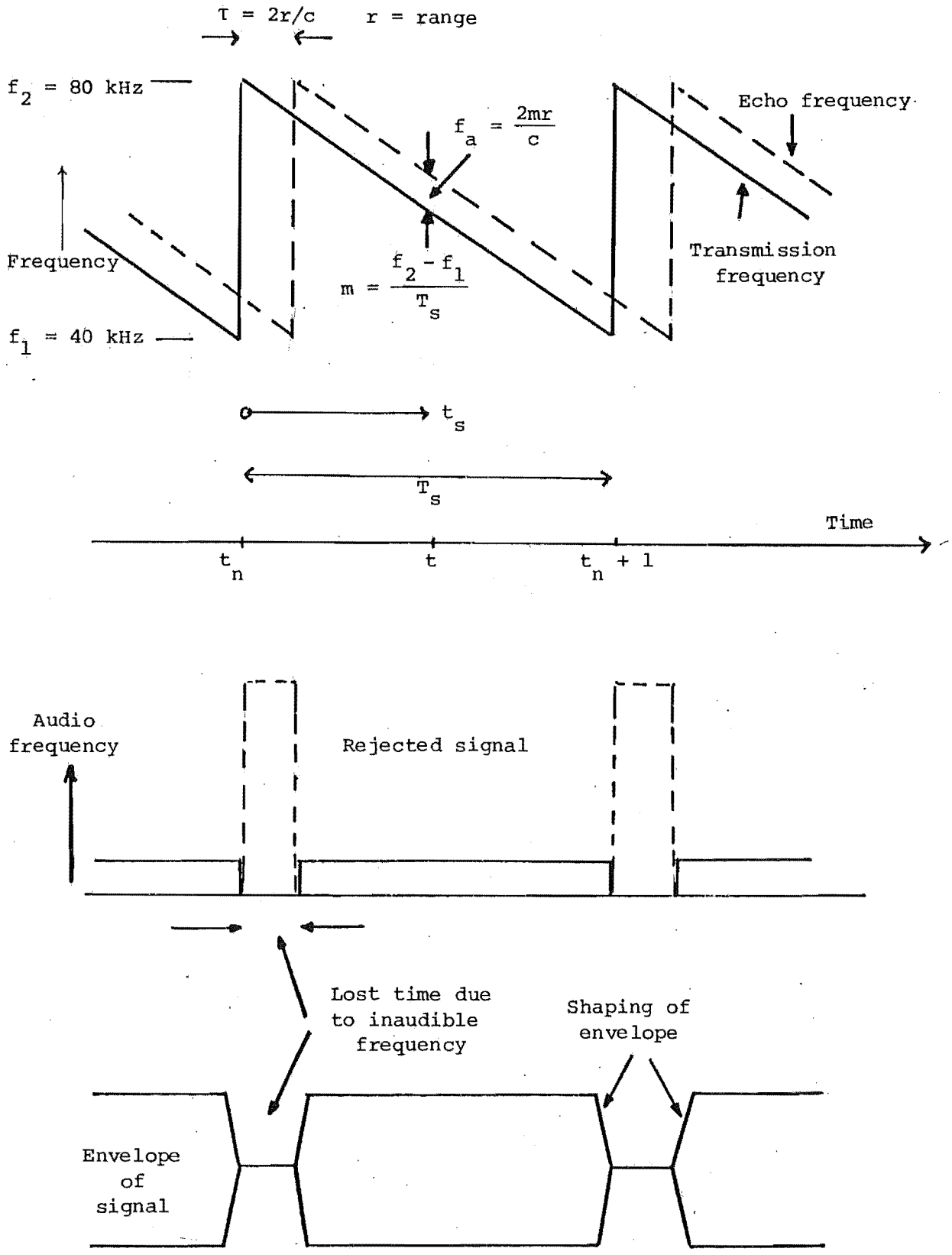


Fig. 5: F.M. sweep parameters and audio output signal (stationary conditions).

where t_s is the time during a sweep cycle, i.e. $0 \leq t_s \leq T_S$

T_S is the sweep period

m is the sweep slope to be chosen.

This signal may be assumed to have a rectangular envelope of duration T_S neglecting the "loss time". Any instant, $t = t_n + t_s$; where t_n is the beginning of the n-th period of the modulating function , and t_s varies between 0 and T_S .

At the time t , the instantaneous transmitted frequency is

$$f_T(t) = f_2 - m\, t_s \qquad (2)$$

Under perfectly stationary conditions, the received signal $S_R(t)$ is a delayed replica of the transmitted signal reflected by an object at distance r . Hence

$$S_R(t) = \alpha(r)\, S_T(t - \tau) \qquad (3)$$

If C is the velocity of sound in the propagating medium then τ , the time delay, is $2r/C$; $\alpha(r)$ is the attenuation due to range r . The instantaneous phase angle of the received signal is then

$$\theta_R(t) = 2\pi\left(f_2\,(t_s - \tau) - \frac{m}{2}\,(t_s - \tau)^2\right) \qquad (4)$$

The received frequency is given by

$$f_R(t) = \frac{1}{2\pi}\,\frac{d\,\theta_R(t)}{d\,t} = f_2 - m\left(t_s - \frac{2\,r}{C}\right) \qquad (5)$$

The range coding frequency of the system is obtained from the difference between the transmitted and received frequencies at time t.

$$f_a = f_R(t) - f_T(t) = \frac{2\,m\,r}{C} \qquad (6)$$

and is made audible by a suitable choice of m .

Thus, whilst the transmitted and received signals are time

varying and inaudible between 40 and 80 kHz, the audible signals at the channels are of the same constant frequency proportional to the distance r, in the interval $t_n < t < t_{n+1}$ (neglecting short transient intervals $<< T_s$). The interaural amplitude difference (I.A.D) at two ears is designed to be proportional to the azimuthal direction of the object, θ.

$$I.A.D = 20 \log \frac{A_R}{A_L} = K\theta \qquad (7)$$

where A_R and A_L are the amplitudes of the audible signals presented at the right and left ears respectively. The constant K depends on the shape of the receiver responses. By a suitable choice of K, the direction code can be matched so that the estimated direction as perceived by the user is equal to the actual direction θ.

The I.A.D code is obtained in practice by using a specially designed transmitter-receiver transducer arrangement [5]. The ideal overall transducer response is shown by Rowell [3] to be

$$A_K(\theta) = C \exp\left[- \frac{(\theta - \alpha)^2}{4 \alpha k} \right] \qquad (8)$$

$$A_L(\theta) = C \exp\left[- \frac{(\theta + \alpha)^2}{4 \alpha k} \right] \qquad (9)$$

where α is the splay angle of the receiver transducers, C a constant, and k is related to the slope of the I.A.D curve, K, by

$$K = \frac{20 \log_{10} e}{k} \qquad (10)$$

$$e = 2.718$$

If an object consists of a number of surfaces, each scattering energy back to the receiver, and the distance to each surface is $r + \Delta r_i$, $i = 1, n$; the audible signals corresponding to the

object will be

$$V_R(t) = \sum_{i=1}^{n} \alpha(r)\, A_R(\theta_i)\, \exp[2\pi j \frac{2m}{C} (r + \Delta r_i)t] \qquad (11)$$

$$V_L(t) = \sum_{i=1}^{n} \alpha(r)\, A_L(\theta_i)\, \exp[2\pi j \frac{2m}{C} (r + \Delta r_i)t] \qquad (12)$$

The perceived sound is a complex tone whose frequency spectrum and diffusion in direction are uniquely related to the geometry of the object in space, and consequently having a unique audible character.

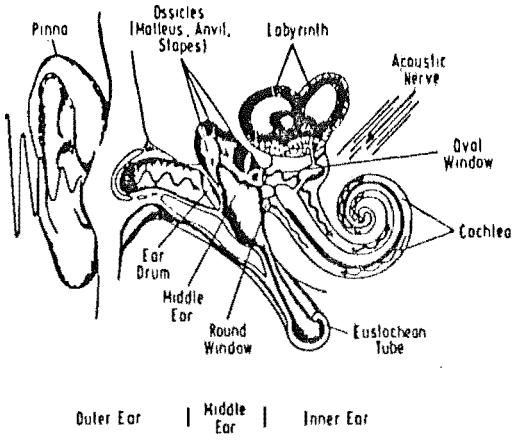
The auditory display of spatial information described so far is only for the perfectly stationary situations. In realistic conditions, there is always continuously varying movement between the environment and the transducers producing Doppler shift in the reflected signals. The audible signals change to patterns of sound variation uniquely related to the relative changes in range and direction of objects. The forms of these patterns are investigated in chapter 3.

Appendix VIII gives a detailed study on the ambiguity function and the resolution performance of the CTFM sonar.

1.4 AUDITORY SYSTEM - AN UNUSUAL FREQUENCY ANALYSER

Since the range of an object is coded in frequency, the resolution and detection performance of the man-machine system, as a whole, is essentially dependent on the capability of the human ear to analyse the spectrum of the sound produced by the sonar.

The first main step of the process of frequency analysis is basically a matter of filtering. In a physical frequency analyser such as a bank of parallel electrical filters with one common input and several different outputs, it is the widths of the filters which determine the frequency resolution of the system, the amount of noise



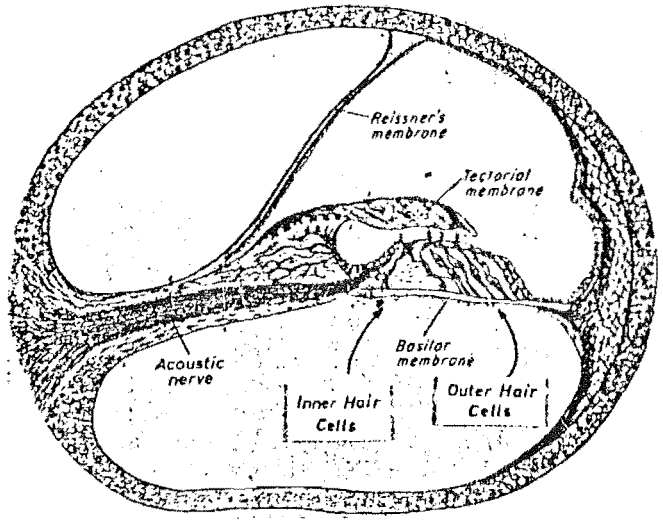
a

Fig. 6: The cochlear analyser.

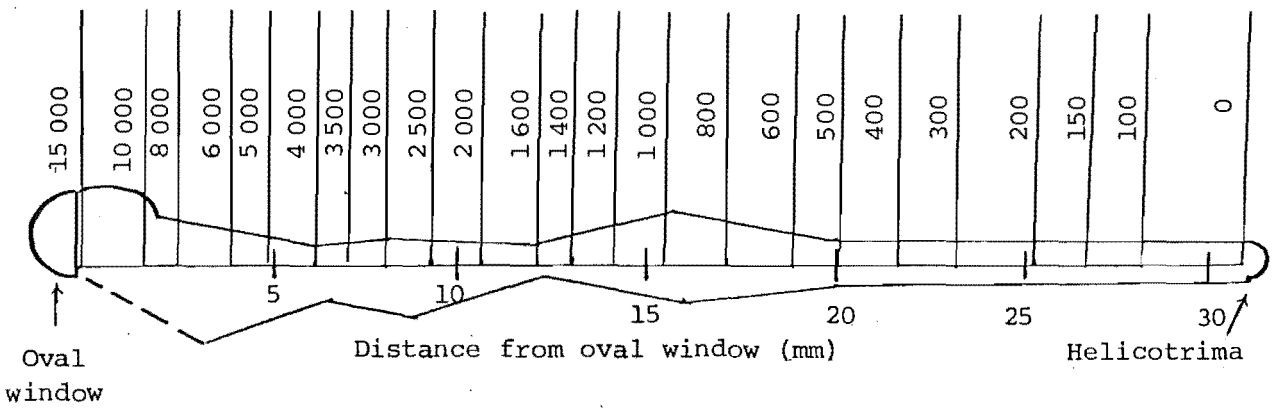
a) Position of the cochlea in human ear.

b) Cross section through cochlea showing fluid-filled canals and basilar membrane supporting hair cells.

c) Characteristic frequency region on basilar membrane.



b



c

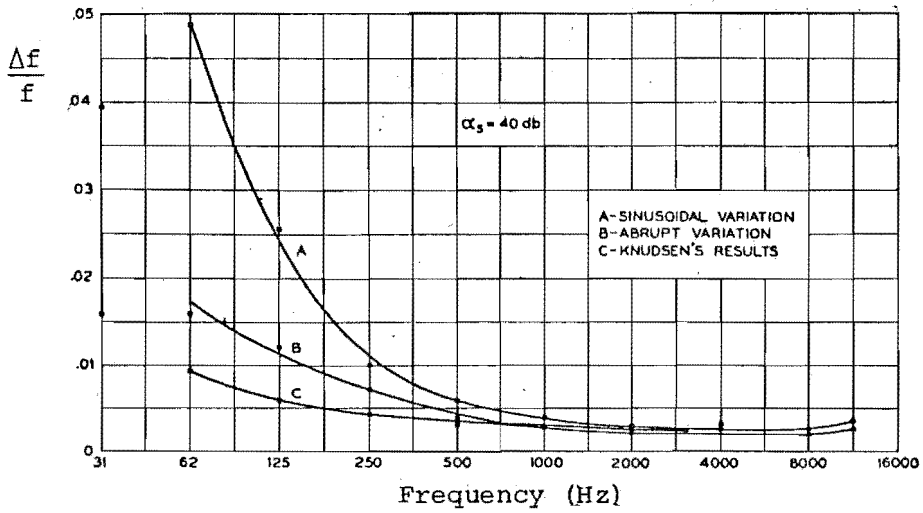
effectively masking a signal and the transient times of the responses. However, psychophysical measurements of the auditory frequency analysis performance have shown that the ear is a very unusual frequency analyser. Some of its basic characters are investigated in this section.

The "cochlea" in the inner ear is the primary frequency selective part of our hearing organ. Inside the cochlea, the "Basilar membrane" supports the "organ of Corti" where the mechanical to neural transduction is affected by the "hair cells" (Fig. 6a,b). Helmholtz visualized the basilar membrane as a succession of tuned strings (as in a piano) resonant at different frequencies [13]. However, it was Békésy [14] who discovered the "travelling waves" on the basilar membrane and recognised the medium of these travelling waves as a "non-uniform" transmission line: high frequencies travel only a short distance on the basilar membrane from the "stapes" and are then attenuated, while low frequencies travel farther along the basilar membrane - the lower the farther - before being stopped. Thus each frequency in the audio frequency range has its own "place" on the basilar membrane where it will cause maximum vibration. This observation has led to the so-called "place theories" of pitch perception, according to which the position of maximum vibration of the basilar membrane determines the pitch of a pure tone (Fig. 6c). The frequency resolution of the basilar membrane as observed by Békésy is very low. He found that two tones, f_1 and f_2, simultaneously sounded must be separated by a fractional difference of $\frac{\Delta f}{f} \simeq 30\%$, where $\Delta f = f_1 - f_2$ and $f = f_1 + f_2/2$, before usual fatigue effects appear. Békésy considered this as the limen where f_1 and f_2 could be heard as distinct tones.

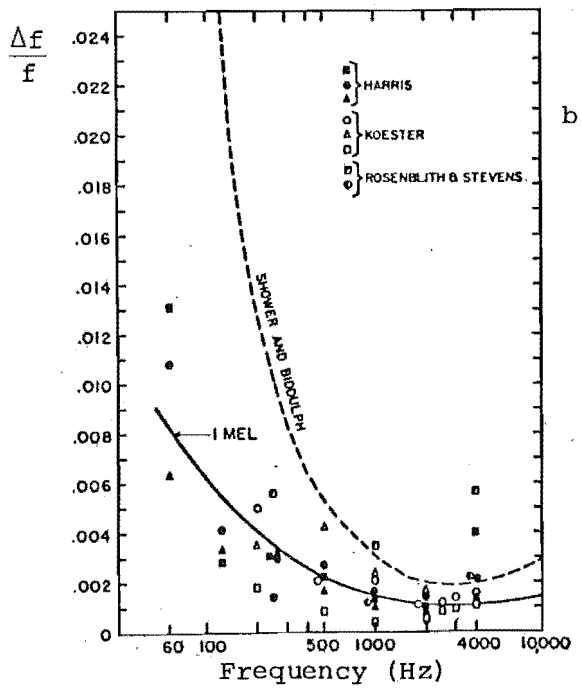
Below this limen, beats are produced; the stimulus magnitude fluctuates continually so that during the silent periods the nerve endings have time to recover. In another experiment using two series of tone pulses of the same amplitude, with each pulse in a series filling the silent period between pulses in the other series, Békésy found that, at 2000 Hz, a frequency difference of 60 Hz was required so that the loudness of two tones could be completely separated.

Several psychoacoustic experiments were carried out to measure the auditory frequency resolution during the last few decades. Different terms, such as frequency discrimination, pitch discrimination, differential sensitivity for pitch, just noticeable frequency difference, minimum perceptible change in frequency, were used to describe the subjective capability in recognising the minimum frequency change with respect to time, either abruptly or in a certain continual variation mode (e.g. sinusoidal modulation, manually adjusting a frequency controlled knob, etc.). The frequency resolution measured by these methods is very high, and the results are relatively close together. They are reproduced in Figure 7 [15,16,17].

Schroeder [18] wrote, "A series obstacle in the place theories was the relatively low frequency resolution (the low Q) of the basilar membrane as observed by Békésy. Psychoacoustically, the just-noticeable frequency difference between two tones presented in succession to a human listener is less than 3 Hz at 1000 Hz! etc. - They explained it by assuming sophisticated neural processing following the crude mechanical filtering action of the basilar membrane. We will not trace their intricate (and probably erroneous) reasoning and simply remark that in more recent work, the mechanical filtering of the Basilar membrane was shown to be much more frequency-selective than found by Békésy.



a) Minimum perceptible change in frequency measured by three different methods. (Reproduced from [17])



b) Different results of pitch discrimination collected and measured by Stevens [18]

Fig. 7: The just noticeable difference in frequency as measured by various methods.

The reason why Békésy did not observe this high frequency resolution is that he worked with dead animals and had to use large sound amplitudes in order to be able to see the basilar membrane motions under his microscope. It is now known, through the work of Rhode [19], that frequency selectivity is substantially lowered at high amplitudes and within minutes after metabolism ceases."

However, the assumption that the mechanical filtering process in the cochlea is the only frequency analysis would not be able to explain the complete difference of the frequency resolution of two tones when presented successively ($\frac{\Delta f}{f} \simeq 0.5\%$) and simultaneously ($\frac{\Delta f}{f} \simeq 40\%$) [3,10,25]. The concept of the critical band as (i) either the limited band of noise effectively masking a tone, or (ii) the frequency band within which the energy of a multiple component tone is summated, etc. was introduced as another approximation of the ear's frequency selectivity, and measured by different methods. Widely diverse results were obtained. Yet, the critical band is still from 20 to 50 times the just-noticeable frequency difference [20,21] (chapter 8). Study on the perception of the sounds containing temporal variations in frequency and amplitude revealed that time varying sounds are more easily resolved and detectable than the stationary pure tones [22,23]. This suggests that central analysers following the basilar membranes filtering exist and are specific for modulated, rather than for constant signals.

1.5 THE NATURE OF THE PROBLEM

Many complex factors are involved in studying the sensory processes and perceptual formation of the user of the multiple object auditory space. To describe and measure the behavioural

characteristics of a binaural sonar user, the science of psychophysics and the sonar system concept must be combined in all investigations. In static situations, each single target is simply described by a tone of more or less single frequency and of a certain interaural amplitude difference, it is of course not very difficult to measure subjective performance. But once the subject and objects are mobile, the sounds become extremely complicated. Each object produces in the auditory space a binaural sound pattern uniquely related to its own motion [24]. It is impossible to handle all the dynamic variables at the same time in the psychophysical experiments. Also the situation becomes even more complicated when objects in the field of view consist of several scattering surfaces. Hence it is necessary to have a detailed investigation on the character of the sound displayed by the binaural sonar in different dynamic situations. From these, essential variables can be selected and simulated in the laboratory experiments. A first approximation to the realistic conditions, simulating moving simple objects, was described in two papers [24,25]. A more detailed investigation is presented in chapters 3 and 4. Experiments to determine discrimination and detection performance of subjects in more realistic and also more complex situations (objects having finite sizes, and auditory space being contaminated by noise) are described in chapters 5, 6 and 8. Chapter 7 studies the acoustic back-scattering characteristics of underwater objects and their reverberant medium, and provides data for the simulations in chapters 6 and 8.

REFERENCES

[1] KAY, L. A new or improved apparatus for furnishing information as to position of objects. Patent Specification No. 978741, The Patent Office, London, 1959.

[2] KAY, L. Blind aid. Patent Specification No. 3366922, United States Patent Office, Washington D.C., 1965.

[3] ROWELL, D. Auditory display of spatial information. Ph.D. Thesis, University of Canterbury, 1970.

[4] MARTIN, G. Electronics and transducers for an ultrasonic blind mobility aid. M.E. Thesis, University of Canterbury, 1969.

[5] KAY, L. A sonar aid to enhance spatial perception of the blind: Engineering design and evaluation. Radio Electron. Eng., Vol.44, 1974, pp.605.

[6] SMITH, R.P. and KAY, L. A fish finding sonar utilizing an audio information display. Digest of Technical Papers, I.E.E.E., Ocean Conf., Panama City, Florida, 1970, pp.113.

[7] SMITH, R.P. Transduction and audible displays for broadband sonar systems. Ph.D. Thesis, University of Canterbury, 1973.

[8] AIRASIAN, P. Evaluation of the binaural sensory aid for the blind. A.F.B. Res. Bull., Vol.26, 1973, pp.51.

[9] KAY, L. Sonic glasses for the blind - Presentation of evaluation data. A.F.B. Res. Bull., Vol.26, 1973, pp.35.

[10] ANKE, von D. Blindenorientierungshielfen mit Ultraschallortungs - Verfahren und horbarer Anzeige. Acustica, Vol.30, 1974, pp.67-80.

[11] KAY, L. An experimental comparison between pulse and frequency-modulation echo ranging systems. J. Brit. I.R.E., Vol.19, 1960, pp.785-796.

[12] KAY, L. Toward objective mobility evaluation, some thoughts on a theory. A.F.B. New York, Monograph 1974.

[13] HELMHOLTZ, H. On the sensation of tone. New York, Dover, 1954.

[14] VON BEKESY, G. Experiments in hearing. New York, McGraw-Hill, 1960.

[15] FLETCHER, H. Speech and hearing in communication. New York, D. Van Nostrand Company Inc., 1953.

[16] STEVENS, S.S. Pitch discrimination, Mels, and Koch's contention. J. Acoust. Soc. Am., Vol.26, 1954, pp.1075-1077.

[17] WEVER, E.G. Theory of hearing. New York, John-Wiley, 1949.

[18] SCHROEDER, M.R. Models of hearing. Proc. I.E.E.E., Vol.63, 1975, pp.1332-1350.

[19] RHODE, W.S. Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. J. Acoust. Soc. Am., Vol.49, 1971, pp.1218-1231.

[20] LICKLIDER, J.C.R. Basic correlates of the auditory stimulus. *In* Handbook of Experimental Psychology, Edited by Stevens, S.S., New York, John Wiley and Sons, Inc., 1951, pp.985-1039.

[21] ZWICKER, E. Scaling. *In* Handbook of Sensory Physiology, Vol.V/2, Auditory System, Edited by Keidel, W.D. and Neff, W.D., 1975, pp.401-448.

[22] WHITFIELD, I.C. and EVANS, E.F. Responses of auditory cortical neurons to stimuli of changing frequency. J. Neurophysiology, Vol.28, 1965, pp.655-672.

[23] KAY, R.H. The hearing of complicated sounds. Endeavour, Vol. XXXV, 1976, pp.104-109.

[24] KAY, L. and DO, M.A. An artificially generated multiple object auditory space for use where vision is impaired. Acustica, Vol.36, 1976, pp.1-8.

[25] DO, M.A. and KAY, L. Resolution in an artificially generated multiple object auditory space using new auditory sensations. Acustica, Vol.36, 1976, pp.9-15.

CHAPTER 2

PHENOMENA RESULTING FROM THE SIMULTANEOUS SOUNDING OF TWO PURE TONES AND THEIR RELATIONS TO THE AUDITORY FREQUENCY RESOLUTION

CHAPTER 2

PHENOMENA RESULTING FROM THE SIMULTANEOUS SOUNDING OF TWO PURE TONES AND THEIR RELATION TO THE AUDITORY FREQUENCY RESOLUTION

2.1 INTRODUCTION

In the preceding chapter, a display of spatial information using audible signals with binaural characteristics was described. According to the description, the audio information obtained from a multiple of objects is extremely complex since a large number of tones, each with its own binaural characteristic, are presented at the same time. In an endeavour toward an understanding of this auditory display, the simplest step is to study the phenomena resulting from the simultaneous sounding of two pure tones so as to determine a meaningful definition of frequency resolution related to spatial resolution. Several different psychological problems are immediately presented under such conditions of stimulation. Indeed, even if the two tones are monaurally presented, depending on the combinations of four variables (two frequencies and two amplitudes) there will be different effects, namely, beats, combination tones (C.T's), masking, fusion, discrimination, etc., on the subject's perception.

In addition to the introduction, there are four sections in this chapter. The first section discusses the investigation of the phenomena due to the simultaneous stimulation of two tones, studied by earlier workers. In the second section, different definitions of frequency discrimination found in literature are examined and a definition for the auditory frequency resolution is proposed. An experiment is then devised to measure this frequency resolution in

the third section. In the last section, the resolution performances in two cases of monaural and binaural perceptions are compared and discussed, and a model is proposed to elucidate the non-linearity of the ear as well as how the C T's can be suppressed at a certain criterion of frequency resolution.

2.2 PHENOMENA RESULTING FROM THE SIMULTANEOUS STIMULATION OF TWO TONES

In 1924, Wegel and Lane [1] of The American Telephone and Telegraph Company designed a very elegant experiment to measure the auditory masking of one pure tone by another. Their experiment was basically to determine the shift of the threshold of hearing of a pure tone when masked by another. However, when using the 80 dB 1200 Hz tone as the masker and increasing the sensation level of the masked tone beyond its threshold of hearing, they could determine the limits in intensity and frequency of the masked tone where either beats, or combination tones, or one tone only, or two tones only etc., would occur. These results were later reproduced by Fletcher [2,3] in an easier understanding form.

Let the primary tone, f_1 , be the masking tone or the masker, and the secondary tone, f_2 , be the masked tone, or the maskee. Figure 1 shows the variations of the threshold shift (i.e. the shift of the threshold of hearing) of the secondary tone as a function of the intensity of the primary tone. These results can be summarized as follows:

(i) When the sensation level of the masking tone is higher than 40 dB, it masks tones of higher frequency more effectively than those of frequency lower than itself. Particularly, a masking tone louder than 60 dB can mask effectively a higher tone of even more

than two octaves apart, but has little effect on a lower tone of about an octave apart.

(ii) When the sensation level of the masking tone is lower than 40 dB, there is little difference between masking a lower tone and a higher tone.

(iii) When the tones are close enough in frequency, the masking curves approach the straight lines with 45° slopes. The tones do not give masking curves in the same sense as when further apart. The curves represent measurements of the minimum perceptible fluctuation of the beating tone.

The results of these measurements were converted to the functions of the frequency of the masked tone in Figure 2 by Fletcher [3]. The frequency of the masking tone is shown on the top of each chart while its intensity is given in the curve itself. Considering the masking curve of the 80 dB 1200 Hz primary tone only. For each frequency of the masked tone, instead of increasing its intensity to the just detectable threshold only, Wegel and Lane increased it up to the level of the primary tone over the range of frequency from 400 to 4000 Hz and determined the criteria where different phenomena due to the simultaneous stimulation of two tones would occur (Figure 3).

A careful analysis was made of the mixture of tones present in the ear when the 80 dB, 1200 Hz primary was present along with the secondary of frequency 700 Hz, and of about the same intensity. The component frequencies were determined by introducing a third tone of known variable frequency and determining the frequencies at which beats occur. The components found in the mixture were:

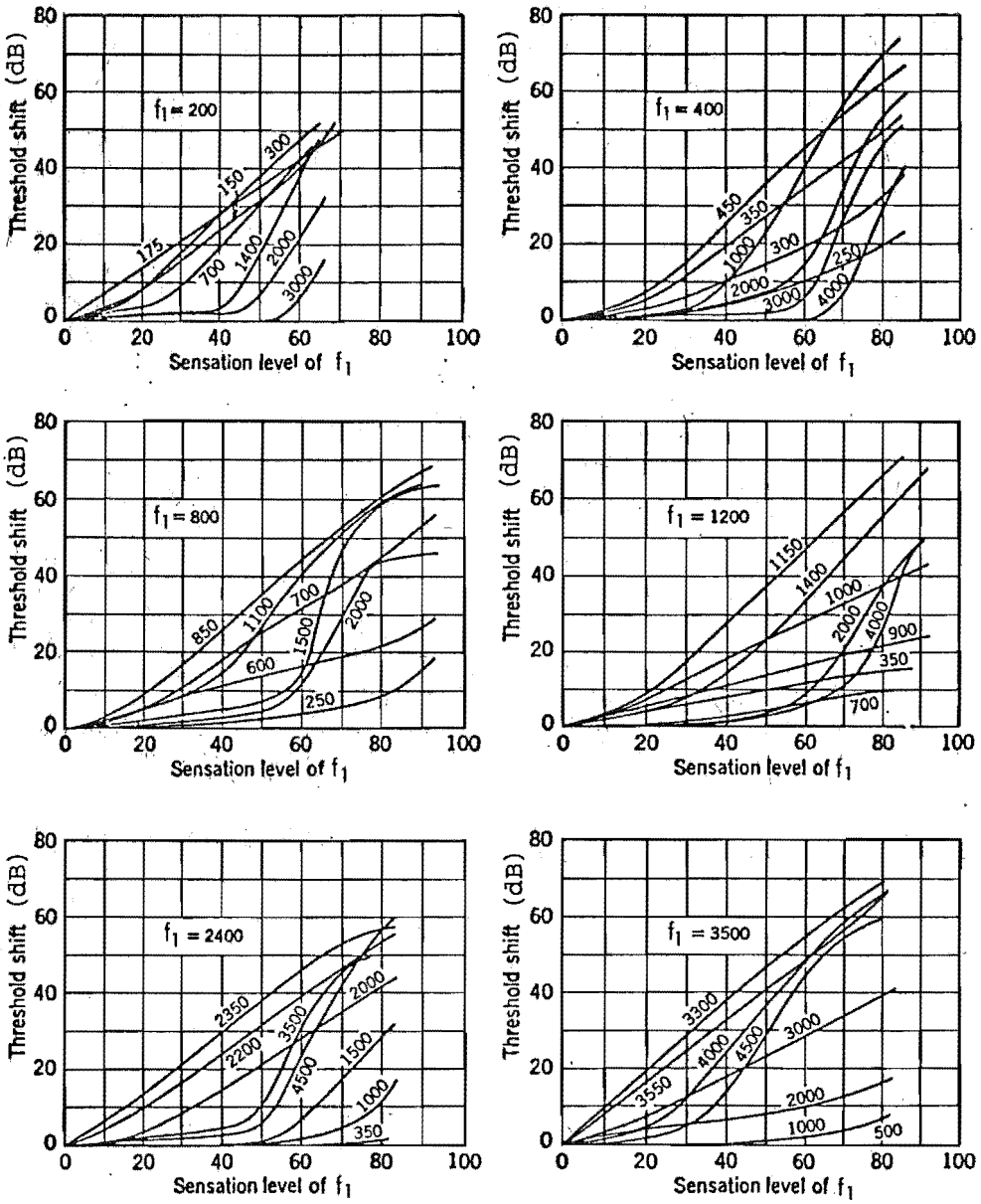


FIG. 1. The masking of secondary by primary components. The primary frequency (in Hz) is given in each plot, and the secondary frequency (in Hz) is attached to the curves. The threshold shift of f_2 and the sensation level of f_1 are scaled in dB. (Ref. 1,2,3)

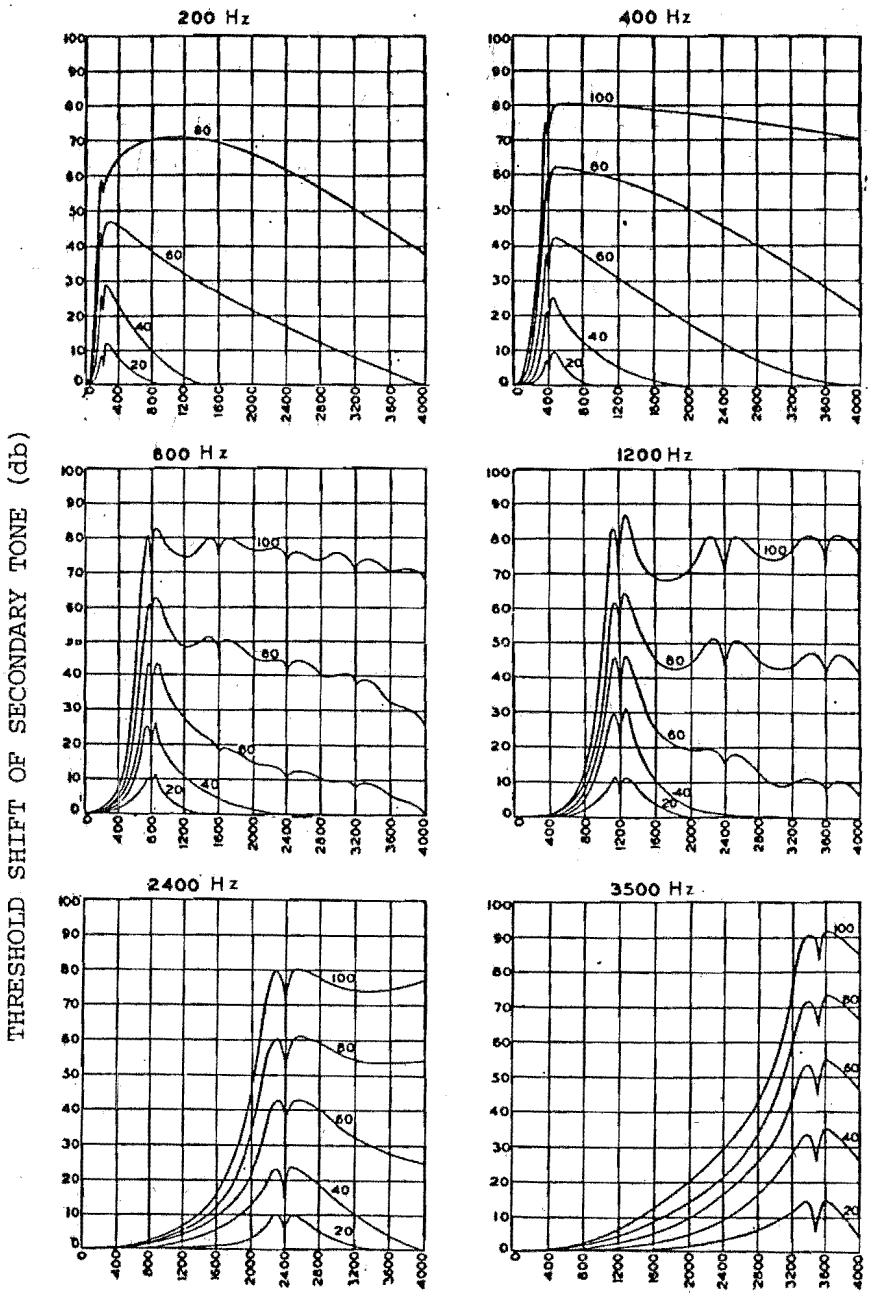


FIG. 2. Variation of the threshold shift of the secondary with respect to its frequency f_2. The frequency of the primary is given at the top of each chart, and its level (in dB) by the number on each curve (Ref. 3).

TABLE 1 - COMPONENTS FOUND IN THE MIXTURE OF SUBJECTIVE TONES

COMPONENT	FREQUENCY (Hz)	COMPONENT	FREQUENCY (Hz)	COMPONENT	FREQUENCY (Hz)
f_1	1200	$f_1 - f_2$	500	$f_1 + f_2$	1900
f_2	700	$2f_1 - f_2$	1700	$2f_1 + f_2$	3100
$2f_1$	2400	$2f_2 - f_1$	200	$2f_2 + f_1$	2600
$2f_2$	1400	$2f_1 - 2f_2$	1000	$2f_1 + 2f_2$	3800
$3f_1$	3600	$3f_1 - f_2$	2900	$3f_1 + f_2$	4300
$3f_2$	2100	$3f_2 - f_1$	900	$3f_2 + f_1$	3300
$4f_2$	2800				

No attempt was made to determine their magnitudes although this can probably be done approximately by measuring the intensity of the exploring tone at which the beats at each frequency are most prominent. Except for the absence of frequency $4f_1$, this series is all that would be expected if the response of the ear were non-linear and represented by the equation

$$x = a_o + a_1 p + a_2 p^2 + a_3 p^3 + a_4 p^4 \qquad (1)$$

where x is the response of the mechanism of the middle ear; a_o, a_1, a_2, etc., are constants, and p is the pressure in the ear canal.

Later, 1929, Wever [4,5] made a detailed investigation into the beat phenomenon and described this by three clearly distinguishable stages, appearing successively as the frequency-difference of two tones is increased from zero:

(i) Noticeable oscillation, or surges of intensity: the rise and fall of the loudness of the tone is prominent so long as

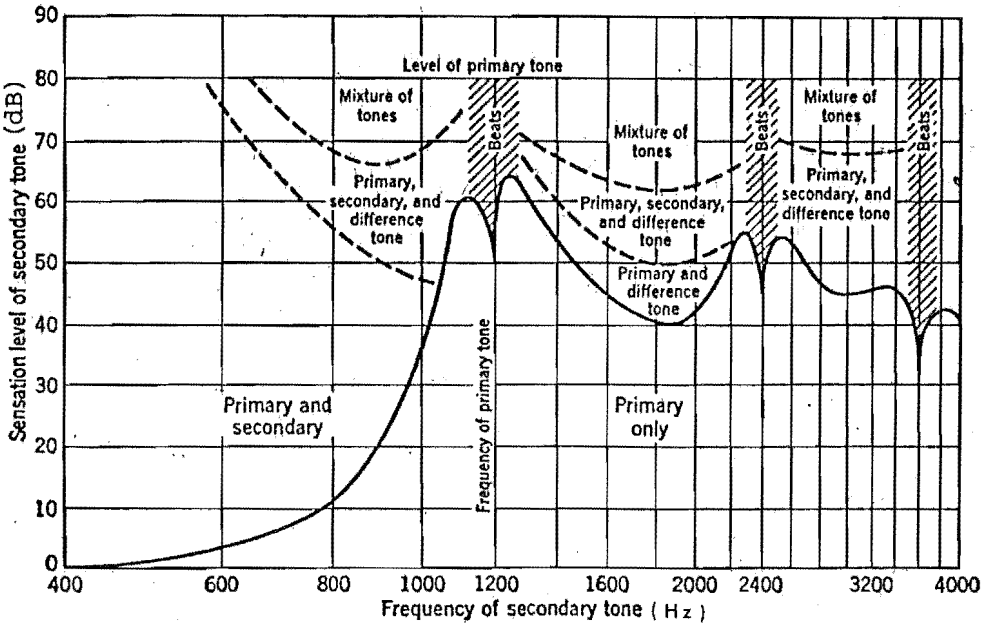


FIG. 3. Criteria of various sensations caused by the simultaneous sounding of two tones. The primary is the 80 db, 1200 Hz tone. The frequency of the secondary varies from 400 to 4000 Hz and its level from 0 to 80 dB (Ref. 1,2,3).

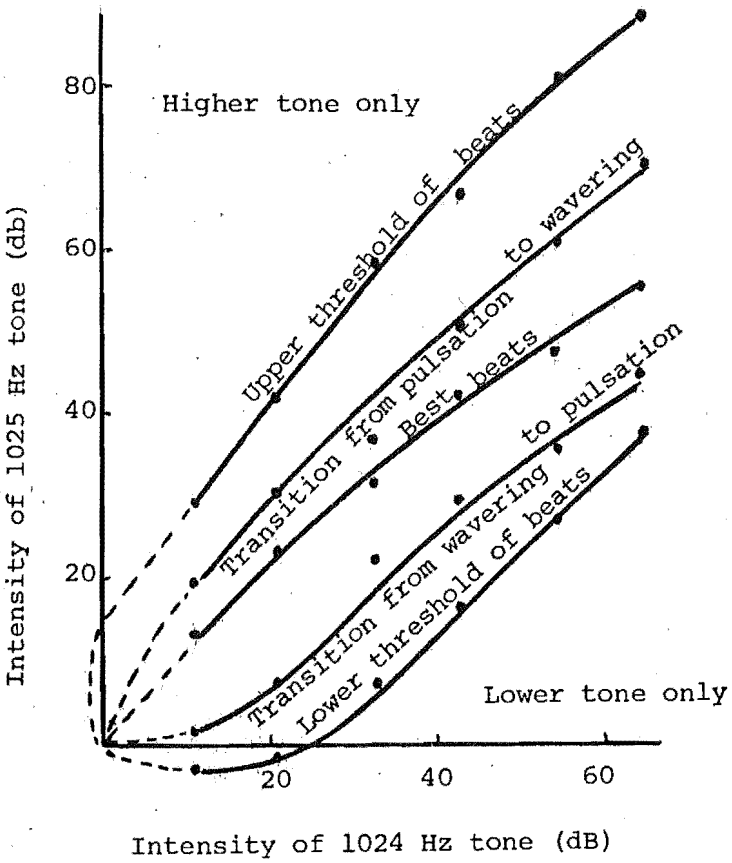


FIG. 4. The effects of intensity on beats (Ref. 4).

the rate is slow.

(ii) Intermittence or pulsation of tone: Subject hears only the pulses of tone separated by silence, when the beats become more rigid.

(iii) Roughness without intermittence: There appears a sort of whir, subject can no longer distinguish individual pulses, and the sound is characteristically rough.

Hence, a definition of the limits within which beats are perceptible must take into account three factors: (1) the pitch-region from which the tones are taken, (2) the intensities at which these tones are sounded, and (3) the particular criterion selected for judgement. Observing two tones equated at an intensity of 20 dB above threshold level, one was at 1024 Hz while the other was varied from that frequency upward, Wever reported that at a rate of 1 beat in 80 sec. the change of intensity was still observable. The three stages of beat are approximately determined by: (i) Oscillation, frequency difference is smaller than 6 Hz; (ii) Intermittence, frequency difference is from 6 to 166 Hz; (iii) Roughness, frequency difference is from 166 Hz to 356 Hz. All these limits are extended as the frequencies of the tones are increased. Investigating the results of very early workers, Wever suggested that the higher values stated for the upper limit of beat perception had been obtained with roughness as the one to judgement, while lower values represented the limits of perception of intermittence.

Using two tones of frequencies 1024 and 1025 Hz, and varying their intensities, the effect of intensity on the thresholds of beats can be illustrated on Figure 4. The phenomenon of beats having been described so far is called the "beats of imperfect unisons". Musicians also experience that if two tones, one the

exact octave of the other, are sounded together, a fused, continuous tone is heard; but if one of the tones is altered slightly in frequency, there will be beats. Likewise, beats arise from other consonant, yet tempered in intervals: fifths, thirds, double-octaves, etc., when they are mistimed. These are the "beats of mistimed consonances". This phenomenon was also shown in Figure 3, and understood as due to the distortion of the tone when conducted through the middle ear. Fortunately, the beats of mistimed consonances are not very prominent, so its effect on the frequency resolution which will be defined later is not significant. Another main phenomenon, namely, Combination tones, which strongly affects the determination of the threshold of auditory resolution will be considered next.

The concept of auditory non-linearity that Helmholtz introduced in the 19th century to account for Combination tones (CT's) is that the auditory process suffers significant percentage of distortion at high sound levels. Little work was found to challenge the concept that auditory mechanical response is linear at low sound levels until 1955, when Zwicker [6] measured the CT's perceived with two-tones stimulus by adding a third tone with the appropriate frequency, phase and amplitude to cancel the C T. His data showed that the C T, having a frequency of twice the lower minus the higher stimulus frequencies, did not increase in percentage distortion with increasing level. Later Plomp [7] reported a much richer spectrum of C T's than the sum and difference tones found by Helmholtz. Using cancellation-tone method, Goldstein and Kiang [8,9], and Hall [10,11,12] carried out elaborate work on measuring the combination tones. The characteristics of the two most prominent C T's; the "cubic difference tone" $2f_1 - f_2$ and the difference tone

$f_2 - f_1$ ($f_2 > f_1$) , can be summarized as follows:

1) The CT, $2f_1 - f_2$, becomes audible at levels near the threshold of hearing where the middle ear is still highly linear.

2) The amplitude of the CT, $2f_1 - f_2$, depends very strongly on the frequency difference of the stimuli, decreasing at a rate of up to 100 dB per octave.

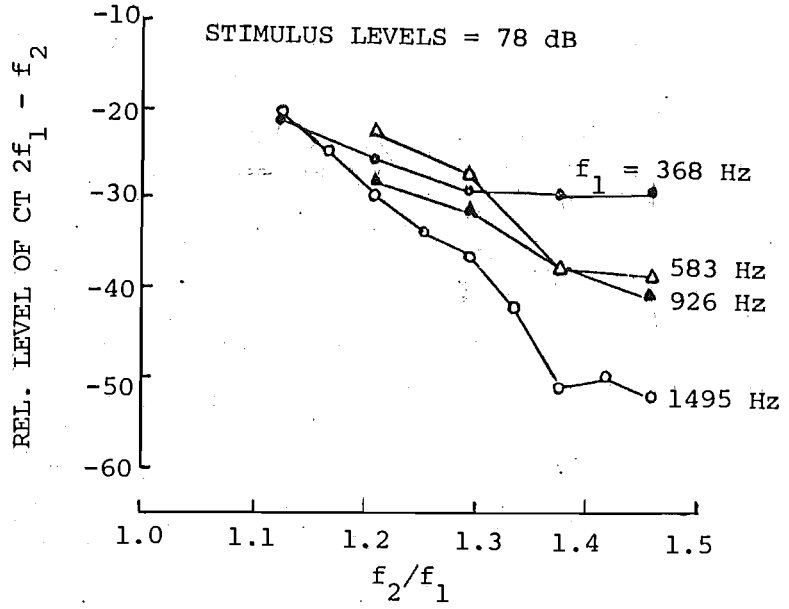
3) The intensity of the $f_2 - f_1$ CT is in some respect similar to that of the $2f_1 - f_2$ CT. It decreases with increasing f_1/f_2, and it changes only slightly with the change of the intensities of f_1 and f_2 .

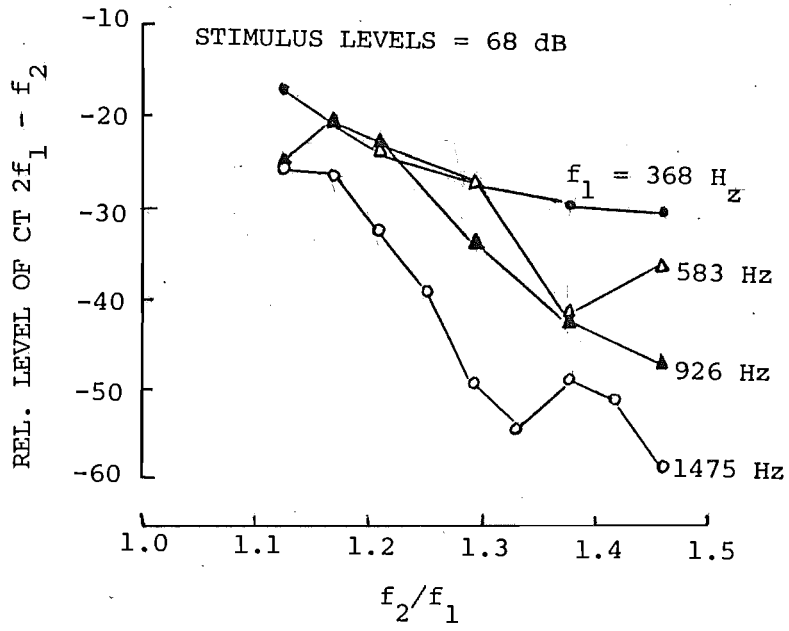
4) A striking difference between the two CT's is that when the stimulus frequencies, f_1 and f_2 , are close together, an increase in the frequencies of the two stimuli decreases the amplitude of the $f_2 - f_1$ CT more than the amplitude of the $2f_1 - f_2$ CT .

The variations of the levels of these CT's (relative to the stimulus levels) with respect to frequency and to intensity are shown in Figures 5 and 6.

2.3 DEFINITION OF FREQUENCY RESOLUTION

The concept of resolution in relation to visual displays of space has been clearly understood for a long time by means of the Rayleigh's Criterion of diffraction theory. In auditory displays, studies on the phenomenon resulting from the simultaneous sounding of two tones shows that if the frequencies of the stimuli are not adequately separated, subjective tones differing from the stimuli can create images of non-existent objects. Therefore the criterion defining auditory frequency resolution must be chosen so that the perception of the tones provide spatial information of the real objects.





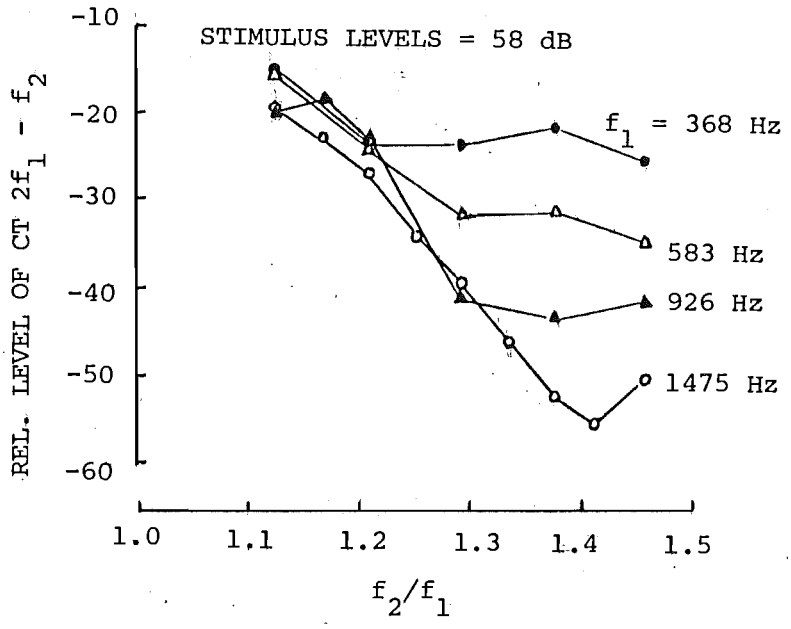


FIG. 5. Variation of the level of CT, $2f_1 - f_2$, with respect to the frequency-difference of two tones, and their levels (Ref.10)

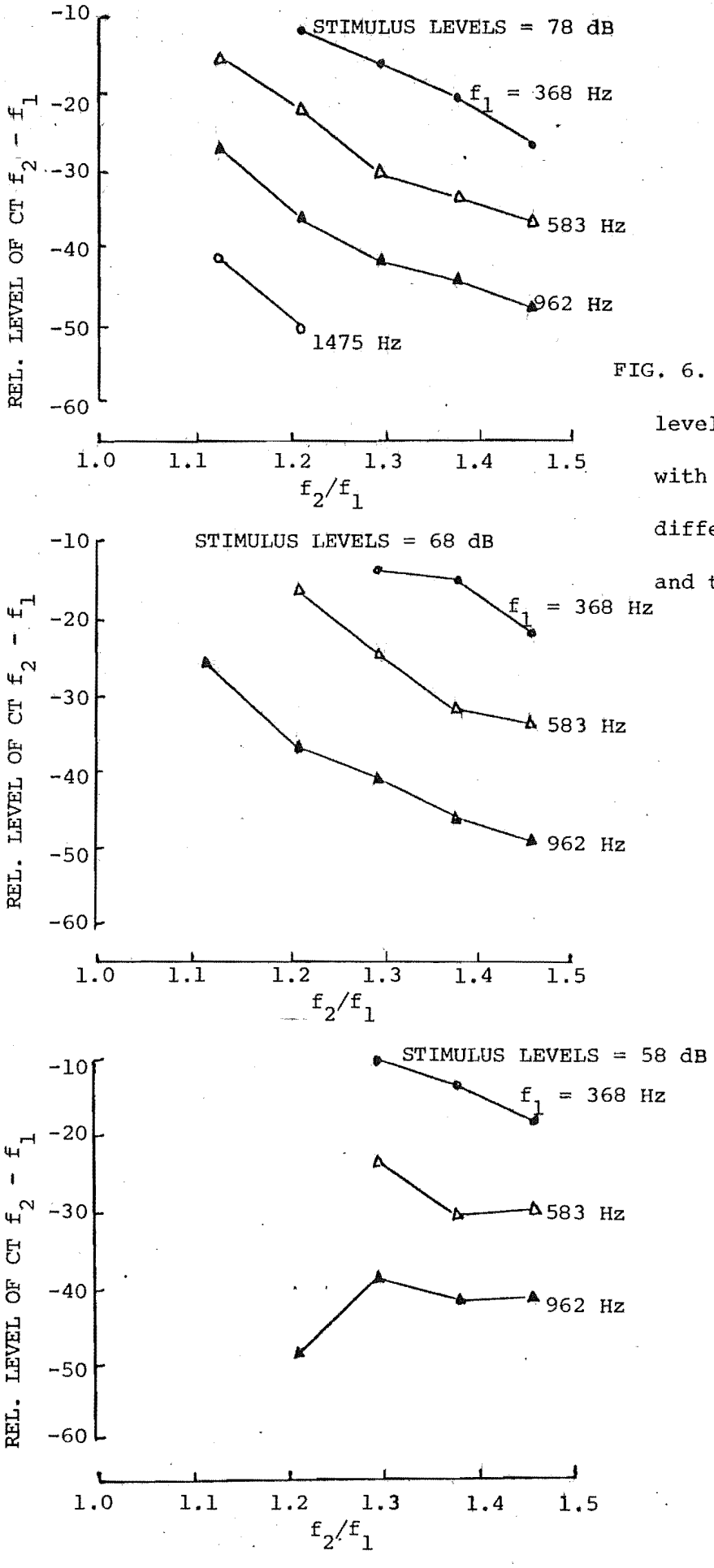


FIG. 6. Variation of the level of CT, $f_2 - f_1$, with respect to frequency-difference of two tones, and their levels (Ref.10)

Observing Figure 3, Wegel and Lane's results show that for the 80 dB 1200 Hz primary tone, the 80 dB secondary tone, when decreasing in frequency from 1200 Hz will meet a criterion where only two tones, the primary and the secondary, are perceived. But it won't meet such a criterion when increasing in frequency, even up to 4000 Hz. The result is very strange and contradictory to the result found by Wever that when the secondary tone is varied upward from the frequency of 1024 Hz of the primary tone, roughness due to interaction between two tones diminishes at a frequency-difference greater than 356 Hz. None of these observations was extended to different frequencies.

Plomp [13], Plomp and Levelt [14], and Plomp and Steeneken [15] attempted to define four different criteria described as follows:

(1) Frequency-difference between partials of a complex of tones required to hear them separately.

(2) Frequency-difference between two simultaneous tones required to distinguish two pitches.

(3) The frequency-difference between two tones for which the harshness of dissonance produced by the tones reaches a maximum.

(4) The frequency-difference for just absence of interference between two tones (beats are inaudible).

These definitions appear to have some meaning related to the auditory resolution. However when investigating the results and the experimental techniques shown in these reports [13,14,15], the writer is convinced that the fourth criterion is equivalent to the end of the second stage of beats described by Wever [4], while the third is actually at the maximum perceptible intermittence of this stage. A forced choice procedure is used in determining the first

and second criteria, where subject is asked to compare a stimulus, which is a twelve-component complex tone in case (1), and is a two-component tone in case (2), with two different single tones, and to choose the single tone which belongs to the stimulus. Obviously the procedure does not testify to either the absence or the existence of the combination tones. So neither of these criteria lead to an understanding of auditory resolution which is applicable to the auditory displays, though they do represent some psychological sensations (Figure 7).

Rowell [16] sought to define the frequency resolution as the least frequency difference between two tones so that one could perceive both clearly and distinctly. Using six subjects he showed that judgements could vary from $\Delta f/f$ being 4% to 40% where Δf is the frequency difference and f the mean of the two frequencies; but only those subjects who gave 40% said they actually heard two tones as distinctly separate. These latter results (Figure 8) were the closest to indicating the resolution capability of the auditory system in terms which are related to the sonar system we are attempting to evaluate.

The purpose of this study is to seek for a definition of the auditory frequency resolution which is related to the spatial range resolution in a sense that when there are two objects presented in the sonic field of the sonar system, the user will hear only two tones and the perceived tones must be the coding tones of the objects (which are the stimuli). Thus the auditory frequency resolution is defined as the least frequency difference between two tones so that they are perceived as being "distinctly separate". The term "distinctly separate" used here implies the absence of interference between the stimulus tones so that they are the only tones perceived

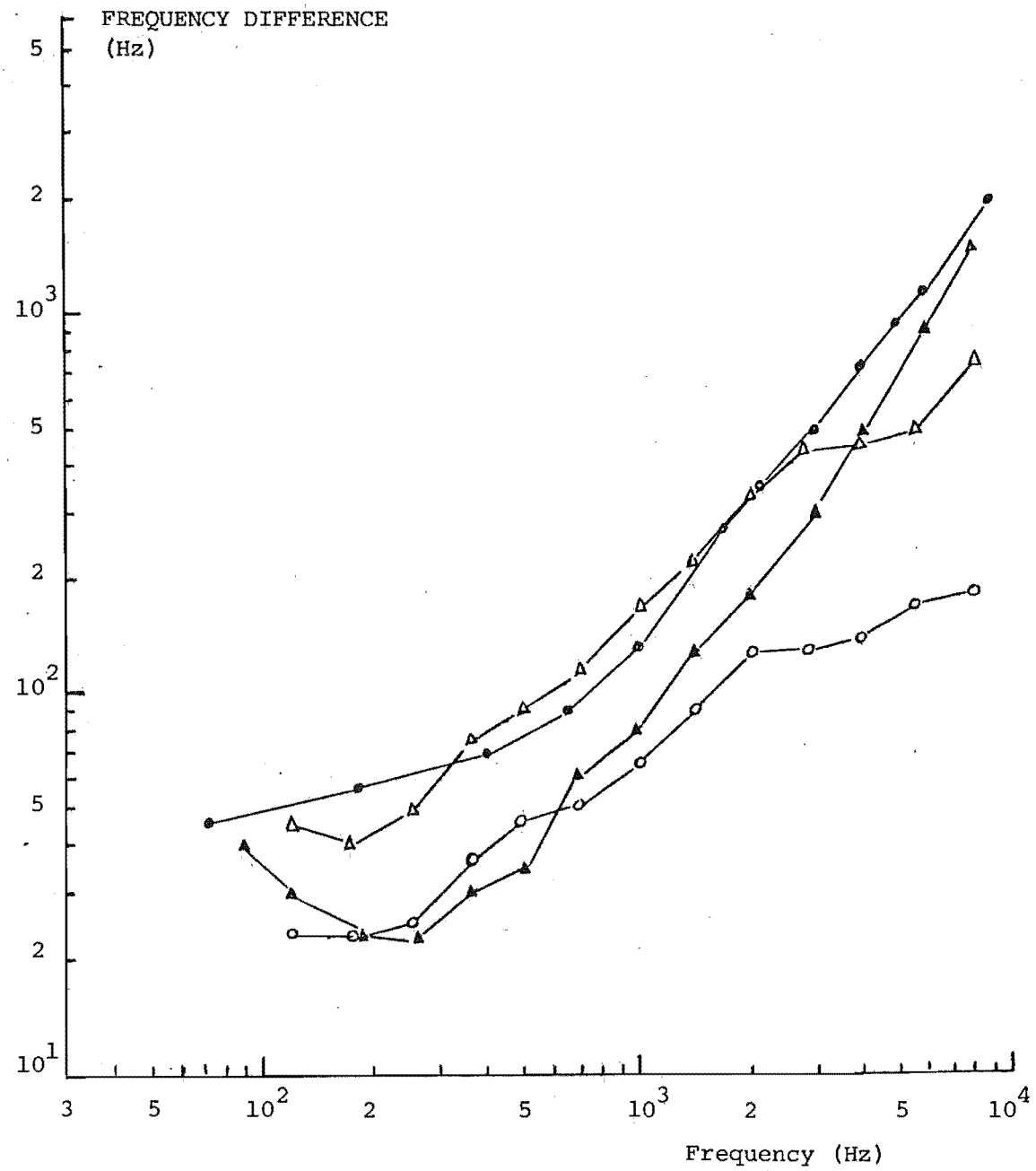


FIG. 7. Four Criteria of Sensations defined by Plomp et al.

● Frequency difference between the partials of a complex tones required to hear them separately.

▲ Frequency difference to distinguish two tones simultaneously presented.

o Frequency difference for maximal roughness of two tones

Δ Frequency difference for just absence of beats given by two tones

(Ref. 13,14,15)

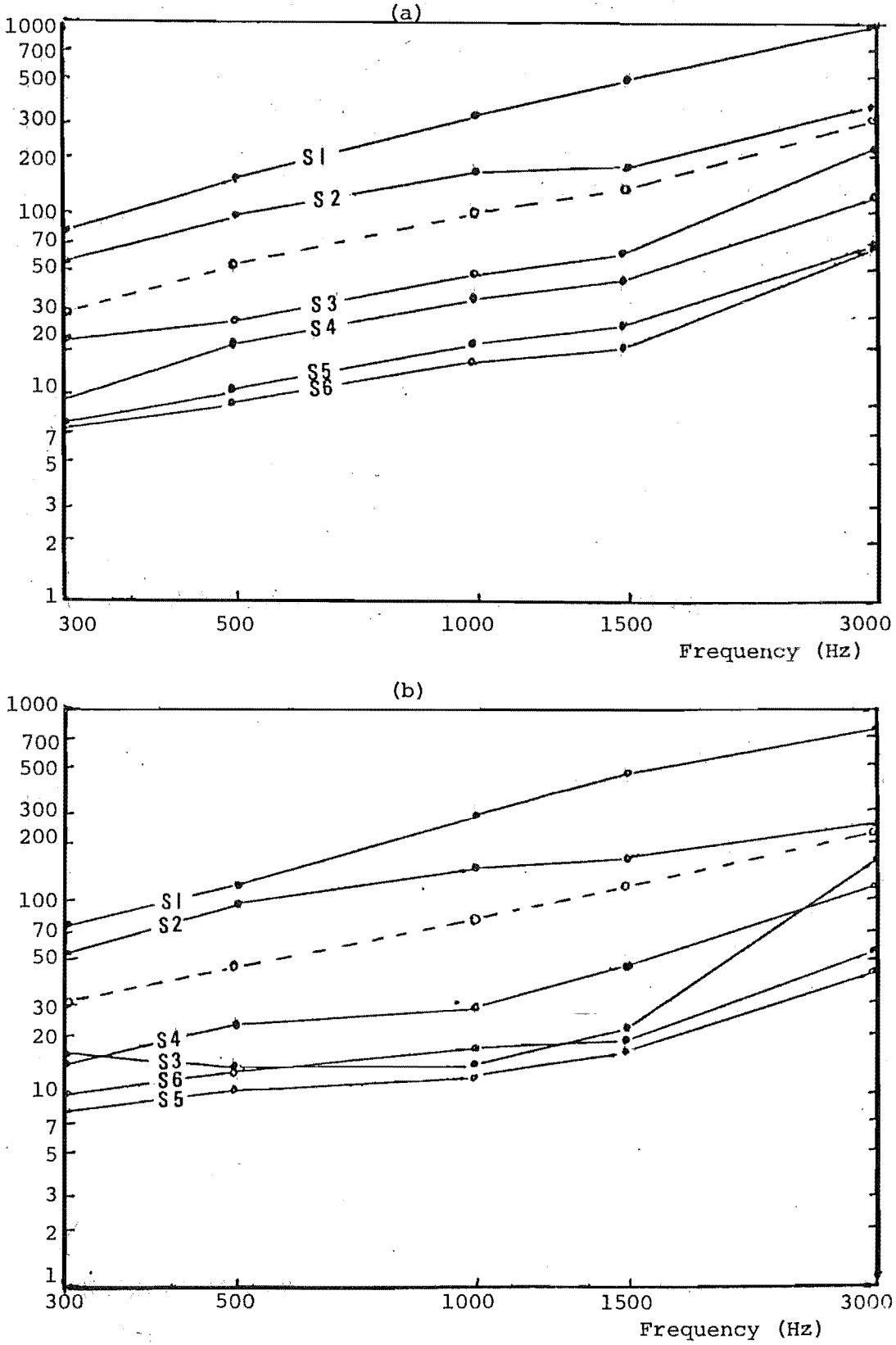


FIG. 8. Frequency Resolution defined by Rowell [16]. Performances of six subjects are significantly different. Dashed curve denotes the average resolution. (a) Pulsed tone (period 330 ms, blanking 30 ms), (b) Continuous tone.

by subject. To determine this frequency resolution at different frequencies an experiment has been carried out. The description is given in the next section.

2.4 MEASUREMENTS OF FREQUENCY RESOLUTION

An experiment is now devised to measure the above defined frequency resolution. We are considering a stimulus of two pure tones, one of fixed frequency and the other of adjustable frequency, whose frequency difference is gradually increased from zero. If the experimental procedure is to simply ask the subject to determine the minimum frequency difference of the two tones so that two components can be heard "distinctly and clearly", highly divergent results will be obtained from different subjects (see Rowell's results, Figure 8). This happens because the above procedure can not verify subjects' judgements. The phenomena resulting from the simultaneous stimulation of two tones are so complicated and subjects may not, perhaps, be sure about the psychological meaning of "distinctly and clearly". It seems that in Rowell's experiment, subjects used different stages of beat phenomenon [4] to judge the resolution of two tones. It is also necessary to note that there are the limits of frequency difference of the stimuli where the C T's can be heard more clearly than the high stimulus tone itself. So when a subject reports that he hears two tones it is not sure that he perceives the stimuli. In order to obtain the right measurement of frequency resolution as defined in the last section, a procedure has been designed to verify the subject's judgement at the criterion of resolution.

(i) TWO and ONLY TWO TONES ARE PERCEIVED by subject, and

(ii) THE PERCEIVED TONES ARE THE STIMULI.

PROCEDURE

Let f_1 be the primary frequency, and f_2 ($< f_1$) be the secondary frequency. Initially f_2 is much smaller than f_1 so that they are distinctly heard as two separate tones. The secondary frequency f_2 is then increased step-by-step until it reaches the maximum value for which both f_1 and f_2 are still resolvable in the sense defined. For each pair of f_1 and f_2, two tests are successively carried out to verify subject's judgement as described in (i) and (ii):

TEST 1 - In this test, the subject's task is to listen to a sound, which randomly includes either two components f_1 and f_2 only, or three components f_1, f_2 and f_3, where the additional component f_3 may have one of the values $2f_2 - f_1$ or $f_1 - f_2$. A "Yes" or "No" answer must be given to the question "is this only TWO TONES?". A "No" answer should be given when three components are included in the stimuli.

TEST 2 - If the answers given to a pair of f_1 and f_2 are correct, test 2 is carried out. Immediately after a "Yes" answer is given in test 1, a single tone is then used as a stimulus through subject's headphones. Either of two questions are then asked "is this the LOW TONE?" or "is this the HIGH TONE?", while the single tone may have the value of either f_1, f_2, $2f_2 - f_1$, or $f_1 - f_2$. If these answers are correct, the tests are repeated with another pair of f_1 and f_2.

The experiment was carried out in an anechoic chamber, and equipment was set up as in Figure 9.

RESULTS

Two types of tones are used in the experiment: (1) Continuous tone, (2) pulsed tone (repetition period 2 sec, blanking duration

200 m secs). The stimuli are of the same intensities, and their total sound level is kept at 70 dB. Results given by four subjects for the continuous tone, and two subjects for the pulsed tone are shown in Table 2 and Table 3. The average value of the frequency difference, Δf , so that two tones can be resolved is present as a function of the Primary frequency (Figure 10). These results have been already shown elsewhere [17].

TABLE 2 - FREQUENCY RESOLUTION OF CONTINUOUS TONES

Δf \ f_1	500	1000	1500	2000	3000	4000
S_1	125	270	400	740	1300	1800
S_2	90	365	480	570	1180	1890
S_3	110	410	615	770	810	1570
S_4	130	370	470	690	1080	1530
Δfav	115	355	490	695	1095	1725

TABLE 3 - FREQUENCY RESOLUTION OF PULSED TONES
(Repetition period 2 s , Blanking 0.2 s)

Δf \ f_1	500	1000	1500	2000	3000	4000
S_1	120	390	530	740	1200	1750
S_3	145	440	685	900	1060	1620
Δfav	135	415	610	820	1130	1750

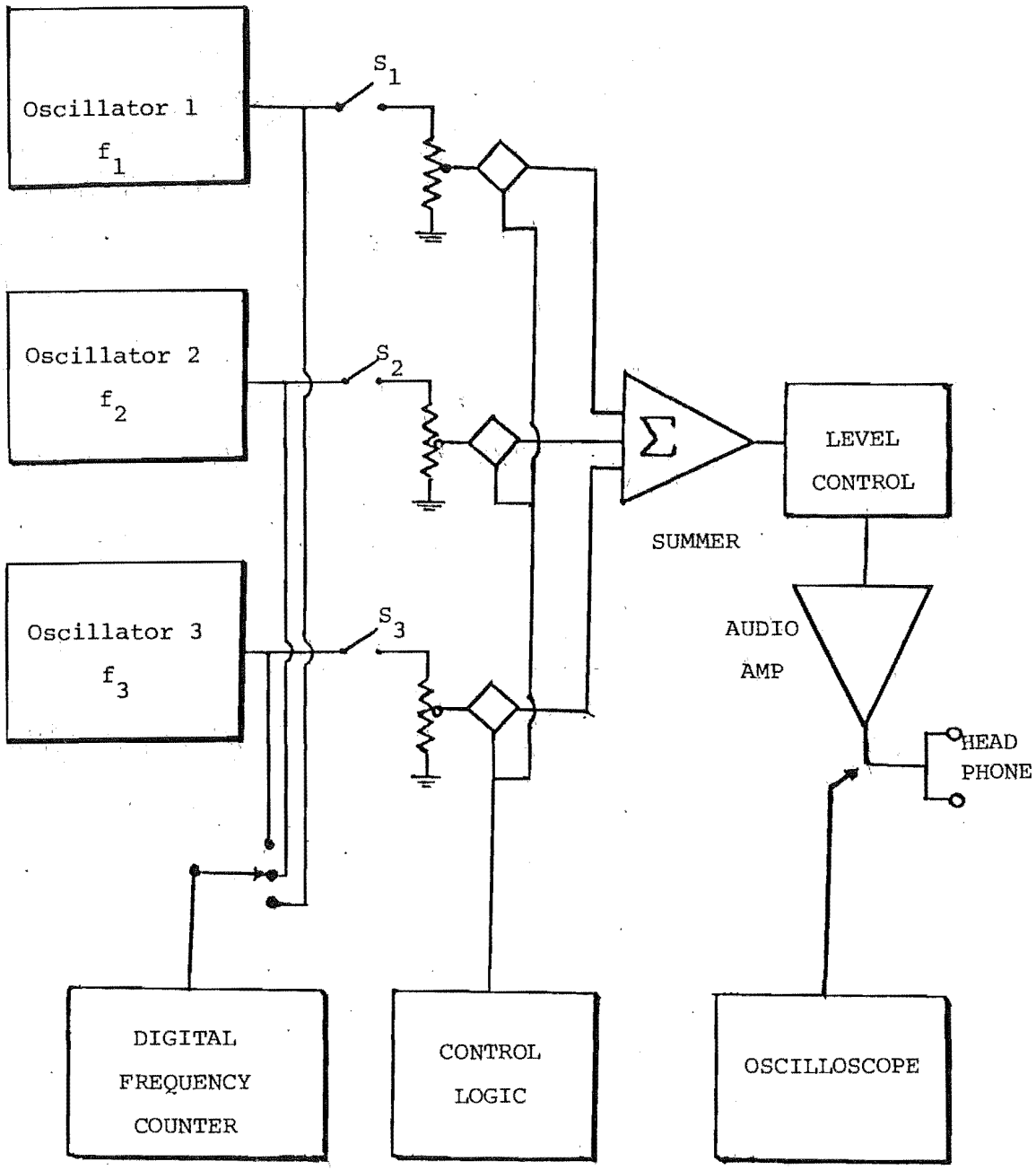


FIG. 9 - Equipment arrangement for measurement of frequency resolution.

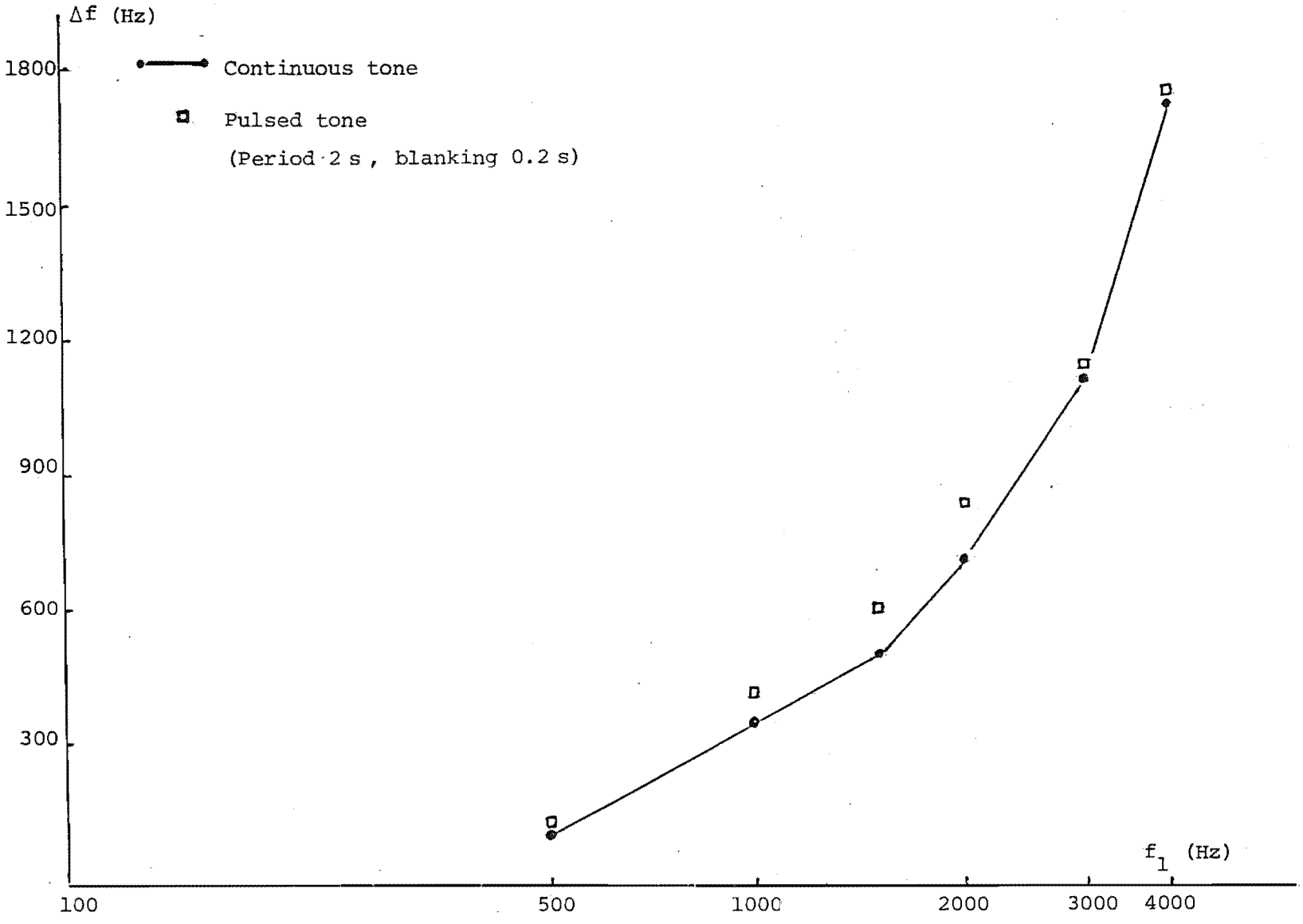


FIG. 10. Mean frequency difference so that two tone can be resolved

2.5 DISCUSSION

From the above results, taking $f = \frac{f_1 + f_2}{2}$ and $\Delta f = f_1 - f_2$, then $\frac{\Delta f}{f}$ is approximately 40%. Rowell [16], in another experiment, used two tones having different I.A.D.'s, and found that subject requires a fractional frequency difference, $\Delta f/f$, of 40% to tell which tone, high or low, is in the left or in the right. The agreement between the results means that when two single tones are really resolvable in frequency, the I.A.D. of each individual tone is perceivable.

Another important point found by Rowell is that the frequency difference at which two tones can be discriminated in direction, left or right, seems not to be dependent on the I.A.D.'s of the tones themselves at all. Little explanation has been found for this fact. Cherry and Sayers [18,19] considered the human auditory system as a cross-correlator and suggested a model for "Binaural Fusion". According to them, for a complex stimulus $F(t) + G(t)$ applied to both ears of a listener, there may exist four possible gestalten described by

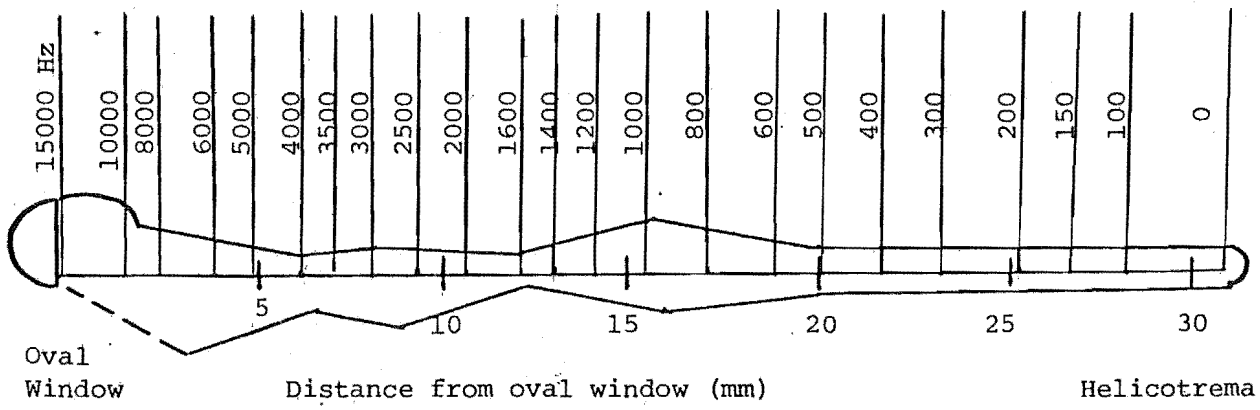
(i) $(F+G)_L : (F+G)_R$; (ii) $F_L : F_R$; (iii) $G_L : G_R$;

(iv) $F_L : F_R$, $G_L : G_R$

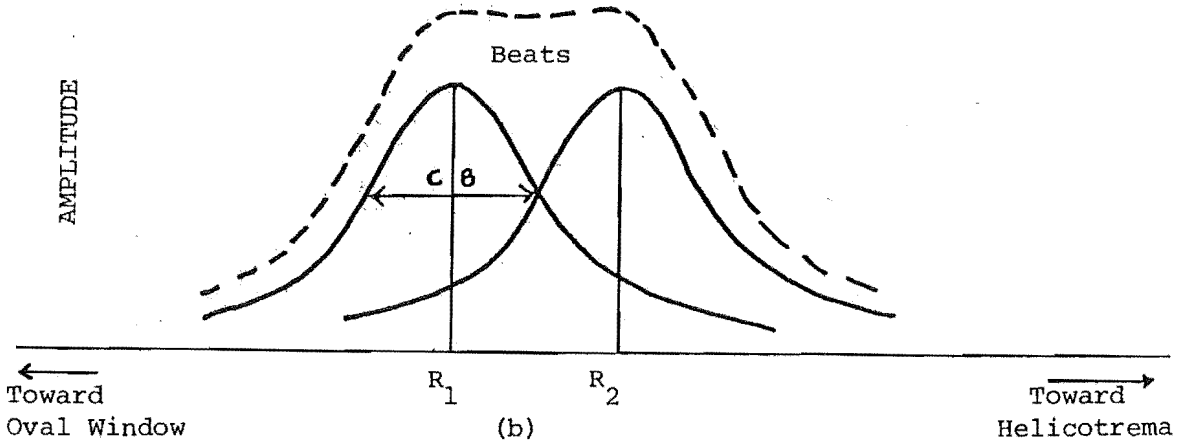
where the suffices L and R indicate left and right ears while ":" indicates interaural fusion. The listener may hear (i) the complex stimulus as a single virtual source, $(F+G)$ at both left and right ears, fused. Alternatively , he may hear (ii) and (iii), one of the single stimuli F or G , in both ears, forming a gestalt, but without attending to the other component. It is possible (iv) that the listener may hear $F_L : F_R$ as one fused source and $G_L : G_R$ as another. Actually the mathematical model of binaural fusion for the

two tone stimulus suggested by Cherry and Sayers mainly explains the complex gestalt (i) which is most readily formed. However, according to that model, if these two tones are not very different in level, the frequencies of the tones will become the dominant variables which determine the formations of the gestalten (ii) and (iii) while the I.A.D.'s of the tones are not important in the process of discrimination.

The results given by the above experiments and Rowell's results clearly show that two tones presented to a subject simultaneously must differ considerably in frequency to be heard as two distinctly separate notes or to be discriminated in direction if there is any difference in the I.A.D.'s of the tones. This frequency difference required for the resolution of tones is about 5 times larger than the critical band width calculated according to Fletcher's measurement of the masking threshold [20], and about 2.5 times the C.B measured by Zwicker [21]. Swets et al's explanation for the difference between the results given for the C.B by assuming that different shapes of the C.B had been postulated [22] seemed to lead to the suggestion that the nerves along the basilar membrane when stimulated by a tone would respond with the amplitudes distributed as a bell shape function having its width equal to that of the C.B. Therefore when two tones are simultaneously sounded, the nerves along the basilar membrane will respond with the amplitudes distributed as two bell shaped spikes only if the frequency difference of the two tones is larger than, say, 2.35 times the C.B; otherwise beats will occur [1] (Figure 11). However, the above single place theory can not explain the existence of the CT $2f_1 - f_2$ ($f_2 > f_1$) when the fractional frequency difference of the two tones is smaller than 40%. Schroeder [23] proposed that the basilar membrane responses to the



(a)



(b)

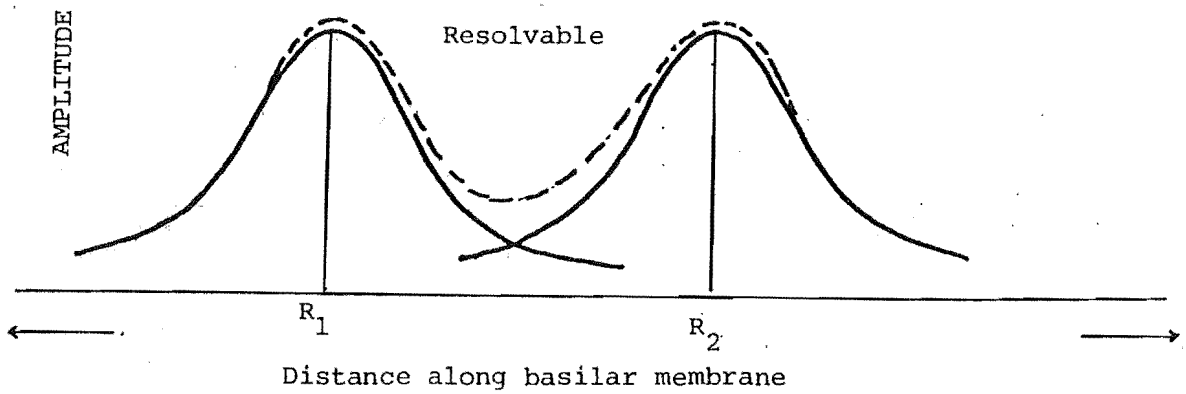


FIG. 11. Hypothetical responses of nerves along basilar membrane (a) Characteristic frequency region on B.M., (b) Beats caused by two tones near in frequency, (c) Resolution of two tones well separated in frequency.

driving pressure of the sound with a velocity following the cube-root law

$$I = \gamma U^{1/3} \qquad (2)$$

where I = the B.M velocity, U = the driving pressure of the sound, and γ = constant,

in order to explain why the amplitude of the cubic difference tone is maximum when the amplitudes of the stimuli are equal [24]. Obviously, Schroeder's model alone can not interpret all the phenomena resulting from the simultaneous sounding of two tones, for example, the strong dependence of the amplitude of the CT, $2f_1 - f_2$, on the frequency difference of the stimuli. A combination of the cube-root law and the place theory, perhaps, elucidates the phenomena better. For example, the cube-root law can be modified to

$$I = \gamma U^{1/3} \otimes h(t) \qquad (3)$$

where h(t) is a place function and the symbol $\otimes$ denotes a convolution process. So in frequency domain, H(f), the fourier transform of h(t), acts as a filter-function. H(f) may have a bell-shape form with a width of a C.B when the stimulus is a single tone. When the stimulus consists of two tones, H(f) will have the combined shape as in Figure 11. Consequently H(f) will filter out, or suppress all the combination tones and beats when the frequency difference of the stimuli is large enough, so that two tones can be perceived distinctly separate.

So far, it has been seen that the non-linearity of the ear results in exceedingly poor frequency resolution. Hence the spatial resolution in the binaural sonar system is obviously unsatisfactory when both the system and the objects are in perfectly stationary conditions. However this rarely happens in realistic conditions.

The movements of either the sonar system or the objects always introduce Doppler shifts in the audio-information, which, as shown in the next chapter, constitute the cues to enhance the auditory frequency resolution significantly.

REFERENCES

[1] WEGEL, R.L. and LANE, C.F. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev., Vol.23, 1924, pp.266-285.

[2] FLETCHER, H. Speech and hearing. N.Y., D. Van Nostrand Company Inc., 1929.

[3] FLETCHER, H. Speech and hearing in communication. N.Y., D. Van Nostrand Company Inc., 1953.

[4] WEVER, E.G. Beats and related phenomena resulting from the simultaneous sounding of two tones - I. Psychological Review, Vol.36, 1929, pp.402-418.

[5] WEVER, E.G. Beats and related phenomena resulting from the simultaneous sounding of two tones - II. Psychological Review, Vol.36, 1929, pp.512-523.

[6] ZWICKER, E. Der ungewöhnliche amplitudengang der nichtlinearen verzerrungen des ohres. Acustica, Vol.5, 1955, pp.67-74.

[7] PLOMP, R. Detectability threshold for combination tones. J. Acoust. Soc. Am., Vol.37, 1965, pp.1110-1123.

[8] GOLDSTEIN, J.L., Auditory nonlinearity. J.Acoust. Soc. Am. Vol.41, 1967, pp.676-688.

[9] GOLDSTEIN, J.L. Neural correlates of the aural combination tone $2f_1-f_2$. Proceedings of IEEE, Vol. 56, 1968, pp.981-992.

[10] HALL, J.L. Auditory distortion products, $f_2 - f_1$, and $2f_1 - f_2$. J. Acoust. Soc. Am., Vol.51, 1972, pp.1863-1871.

[11] HALL, J.L. Monaural phase effect : cancellation and reinforcement of distortion products $f_2 - f_1$ and $2f_1 - f_2$. J. Acoust. Soc. Am., Vol.51, 1972. pp.1872-1881.

[12] HALL, J.L. Monaural phase effects for two-tone signals. J.Acoust. Soc. Am., Vol.51, 1972, pp.1882-1884.

[13] PLOMP, R. The ear as a frequency analyser. J.Acoust. Soc. Am., Vol.36, 1964, pp.1628-1636.

[14] PLOMP, R. and LEVELT, W.J.M. Tonal Consance and critical bandwidth. J.Acoust. Soc. Am., Vol.38, 1965, pp.548-559.

[15] PLOMP, R. and STEENEKEN, H.J.M. Interference between two simple tones. J.Acoust. Soc. Am. Vol.43, 1968, pp.883-884.

[16] ROWELL, D. Auditory display of spatial information. Thesis, Ph.D., University of Canterbury, Christchurch, New Zealand, 1970.

[17] DO, M.A. and KAY, L. Resolution in an artificially generated multiple object auditory space using new auditory senses. Acustica, Vol.36, 1976-1977, pp.9-15.

[18] CHERRY, E.C. and SAYERS, B.McA. Human 'cross-correlator' - a technique for measuring certain parameters of speech perception. J.Acoust. Soc. Am., Vol.28, 1956, pp.889-895.

[19] SAYERS, B.McA. and CHERRY, E.C. Mechanism of binaural fusion in hearing of speech. J.Acoust. Soc. Am., Vol.29, 1957, pp.973-987.

[20] FLETCHER, H. Auditory patterns. Reviews of Modern Physics, Vol.12, 1940, pp.47-65.

[21] ZWICKER, E., FLATTORP, G., and STEVENS, S.S. Critical bandwidth in loudness summation. J.Acoust.Soc.Am., Vol.29, 1957, pp.548-557.

[22] SWETS, J.A., GREEN, D.M. and TANNER JR., W.P. On the width of critical bands. J.Acoust.Soc.Am., Vol.34, 1962, pp.108-113.

[23] SCHROEDER. M.R. Models of hearing. Proceeding of the IEEE, Vol.63, 1975. pp.1332-1350.

[24] ZWICKER, E., Der kubische differenzton und de Erregung des Gehörs. Acustica, Vol.20, 1968. pp.206-209.

CHAPTER 3

ROLE OF DOPPLER EFFECT IN THE BINAURAL DISCRIMINATION OF AUDIBLE FREQUENCY PATTERNS PRODUCED BY WIDE BAND CTFM SONAR

3.1 INTRODUCTION

In Chapter 1, an artificially generated multiple object auditory space for use either by blind persons or in fish location was described. To gather spatial information, a very wide band wide beam CTFM sonar with binaural display was employed. The range of an object (i.e. the distance to an object) was to be coded in the form of rising pitch with increasing range (frequency proportional to distance), and the interaural amplitude difference (IAD) would indicate the azimuthal direction of an object.

The means for producing this form of auditory space was built in the form of a sensory aid for the blind, and a fishing sonar [1 - 5]. Both systems were evaluated during 1970-72. The results of the evaluations showed that human operators, after being trained, could use this auditory display with comparative ease, even though the input to the ears was rich in spatial information. After 40 - 60 hours of training over a period of four weeks, many blind users of the sensory aid were able to travel in a busy pedestrian area with grace and confidence not seen in totally blind people before [6,7,8]. A few, at least, approached the behaviour pattern not unlike that of sighted pedestrians (as recorded on film). An untrained fishing-sonar operator could easily steer the boat over the top of the shoal of fish so as to obtain confirmatory evidence on an echo sounder [5]. When two shoals were located

at the same time in different directions and at different distances, it was reported that the choice could be made to steer over either the larger or the near shoal.

These reports suggest that human operators have learned to use the sound patterns, which flow left or right as objects are passed, rather than the simple range indicating tones of constant frequencies as described for a static situation, since the capability of resolution of two pure tones, as shown in Chapter 2, is very poor. This chapter investigates the Doppler effect in a CTFM sonar, then discusses the capability of discriminating the sound patterns produced by the system under various moving conditions of human operator and objects.

3.2 THE DOPPLER EFFECT IN WIDE BAND CTFM SONAR

The basic parameters of the auditory display of spatial information described in Chapter 1 are only for the stationary conditions - the range of an object in the system field of view, of say 60^{o}, is proportional to the frequency of the audible sound of the device, (typically the sensory aid for the blind has a range code of 941 Hz/m , and the fishing sonar has four different range codes of 53.33, 26.67, 13.33, and 6.67 Hz/m for four different maximum operating ranges), and the direction of an object is indicated by the binaural difference in the loudness of the sounds fed to each ear (typical IAD is from 0.4 to 0.5 dB per degree in the sensory aid, and about 1.5 dB/degree in the fishing sonar [5]. Both systems operate with a frequency band of approximately 40 kHz to 80 kHz.

Under dynamic conditions, the range of an object is no longer simply proportional to the frequency of the displayed sound as

described above. Relative motion between target and sonar system affects the frequency of the sonic wave, changing the frequency f to $f(1 - \frac{2\dot{r}}{c})$, where $\dot{r}$ denotes the range rate and c is the sound velocity. (This effect is called Doppler effect, and the frequency change $\frac{2\dot{r}}{c} f$ is called the Doppler shift). Consequently in a wide band CTFM signal, when the frequency is modulated, say, from 40 kHz to 80 kHz, the higher frequencies undergo larger Doppler shifts than the lower frequencies, and this results in a change of the sweep slope of the received frequency. The effect of this change of the sweep slope of the received frequency on the frequency of the displayed sound (the range coding frequency) is going to be studied for various dynamic situations in this section.

3.2.1 Constant Velocity Situation

If the relative velocity between an object and the sonar system is assumed to be constant (v), the range of the object is given by

$$r(t) = r(o) - vt \tag{1}$$

where $r(o)$ is the initial range at $t_o = 0$, and a radial approach velocity is assumed.

The transmitted signal is cyclic and of the form:

$$S_T(t) = A\exp\left[2\pi j \left(f_2 t_s - \frac{m}{2} t_s^2\right)\right] \tag{2}$$

where, as defined in Chapter 1,

f_2 = upper limit frequency

m = sweep slope

T_s = repetition period

$t = t_n + t_s$ at any instant, where t_n is the beginning of the n-th cycle of the modulating function, and t_s varies between zero and T_s.

A = constant amplitude of the transmitted signal.

The instantaneous transmitted frequency is

$$f_T(t) = f_2 - m t_s, \quad 0 \leq t_s \leq T_s \tag{3}$$

The received signal is then a Doppler shifted "replica" of the transmitted signal $\tau(t)$ seconds ago, which was reflected from the target at distance $r(t_r)$, where $t_r = t - \frac{\tau(t)}{2}$. Thus

$$r(t_r) = r(t) + \frac{v\tau(t)}{2} \tag{4}$$

$$\text{since} \quad \frac{\tau(t)}{2} = \frac{r(t_r)}{c} \tag{5}$$

Solving equations (4) and (5), we have

$$\tau(t) = \frac{2r(t)}{c - V} = \frac{2(r(0) - vt)}{c - v} \tag{6}$$

The received signal is

$$S_R(t) = \alpha(r) \quad S_T(t - \tau(t)) \tag{7}$$

where $\alpha(r)$ is the attenuation due to range r.

The instantaneous phase angle of the received signal is then

$$\theta_R(t) = 2\pi[(f_2 (t_s - \tau(t) - \frac{m}{2} (t_s - \tau(t))^2] \tag{8}$$

The received frequency is given by

$$f_R(t) = \frac{1}{2\pi} \frac{d\,\theta(t)}{dt}$$

$$f_R(t) = f_2 - f_2 \dot{\tau}(t) - m(1 - \dot{\tau}(t))(t_s - \tau(t)) \tag{9}$$

and the range coding frequency is the difference between $f_R(t)$ and $f_T(t)$.

$$f_a(t) = m\tau(t)(1 - \dot{\tau}(t)) - \dot{\tau}(t)(f_2 - m t_s) \tag{10}$$

While the fishing sonar is being operated, the operator steering the boat does not travel at a velocity higher than 3 m/s (about six knots), hence the fractional velocity $\frac{2v}{c}$ is smaller than 4×10^{-3}, where $c = 1500$ m/s is the sound velocity in water.

A blind user of the sensory aid may walk at a speed, at most, 5 km per hour. A sound velocity in air is approximately 340 m/s , hence $\frac{2v}{c} < 8.5 \times 10^{-3}$. Therefore when substituting equation (6) in equation (10), it is reasonable to assume that

$$|\dot{\tau}(t)| \simeq \frac{2v}{c} << 1 \qquad (11)$$

and $$\tau(t) \simeq \frac{2r(t)}{c} \qquad (12)$$

This results in

$$f_a(t) \simeq \frac{2m}{c}(r(0) - ct) + \frac{2v}{c}(f_2 - mt_s) \qquad (13)$$

or $$f_a(t) \simeq \frac{2m}{c} r(t) + \frac{2v}{c} f_T(t) \qquad (14)$$

$\frac{2m}{c} r(t)$ is the static range coding frequency, and $\frac{2v}{c} f_T(t)$ represents the Doppler shift. Figure 1 illustrates how the Doppler shift affects the range coding frequency. To appreciate the significance of this, an example is shown plotted in Figure 2. There is an average shift of $\frac{2v}{c}(f_2 - \frac{m T_s}{2})$ in the range coding frequency, $f_a(t)$, about which $f_a(t)$ varies in a saw tooth cyclic manner. This pattern is due both to the Doppler shift of frequency in the medium and the wide variation in transmitted frequency.

3.2.2 Varying Velocity Situation

A simple constant velocity situation is rare under conditions of human locomotion - the human body is capable of considerable acceleration, and varying velocity could be considered to be more a normal condition generated by man's peripatetic movement. Even for the less complicated case of the fishing sonar, when the boat moves at a constant speed, other movements like the lateral shift of the boat, the unpredictable migration of a fish shoal, and the variation of direction of the boat in the tracking task always vary the relative range rate of the target

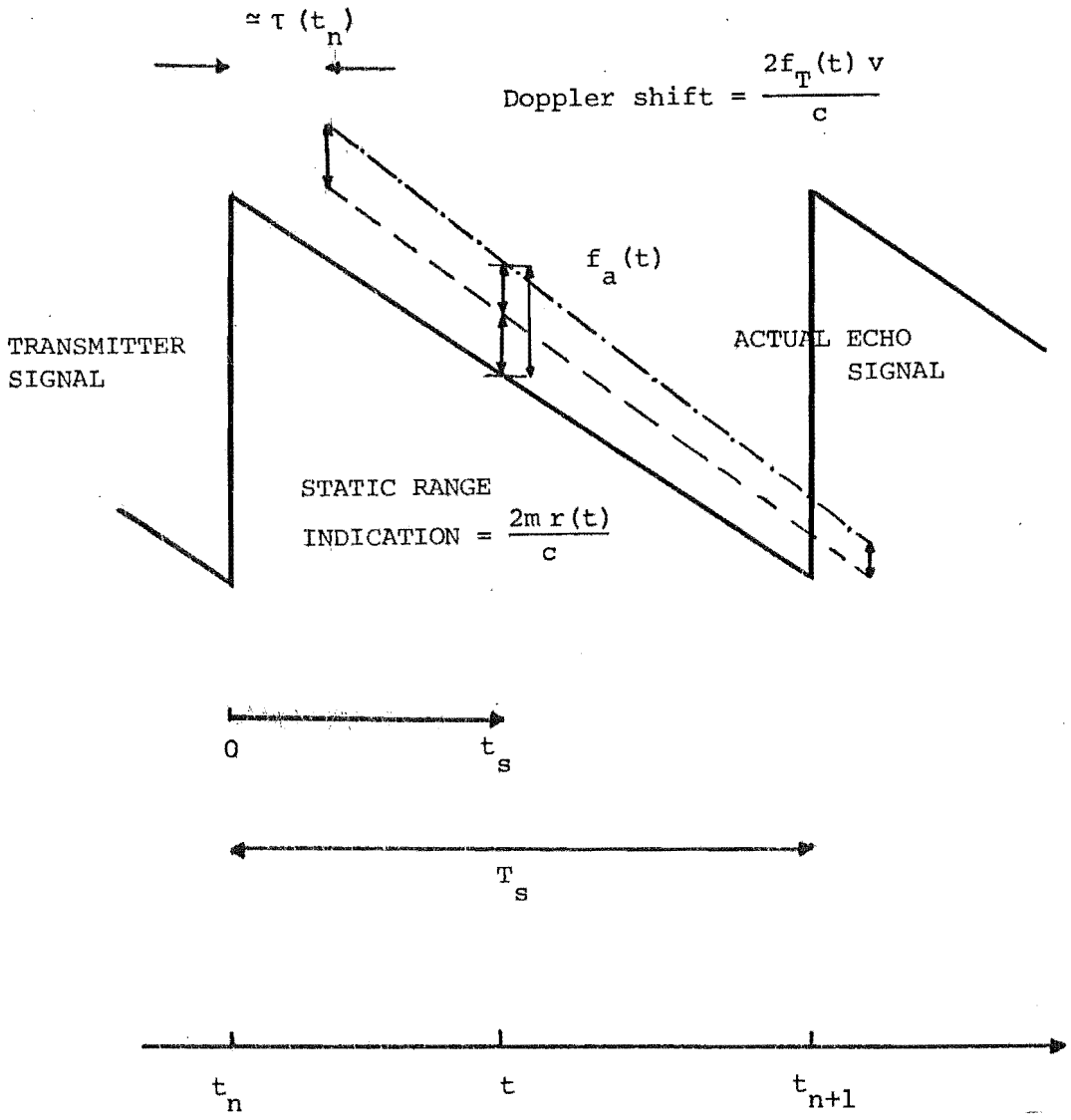


Fig. 1. Doppler Shifted "Replica" of Transmission Wave

(Constant approach velocity $\dot{r}(t) = -v$)

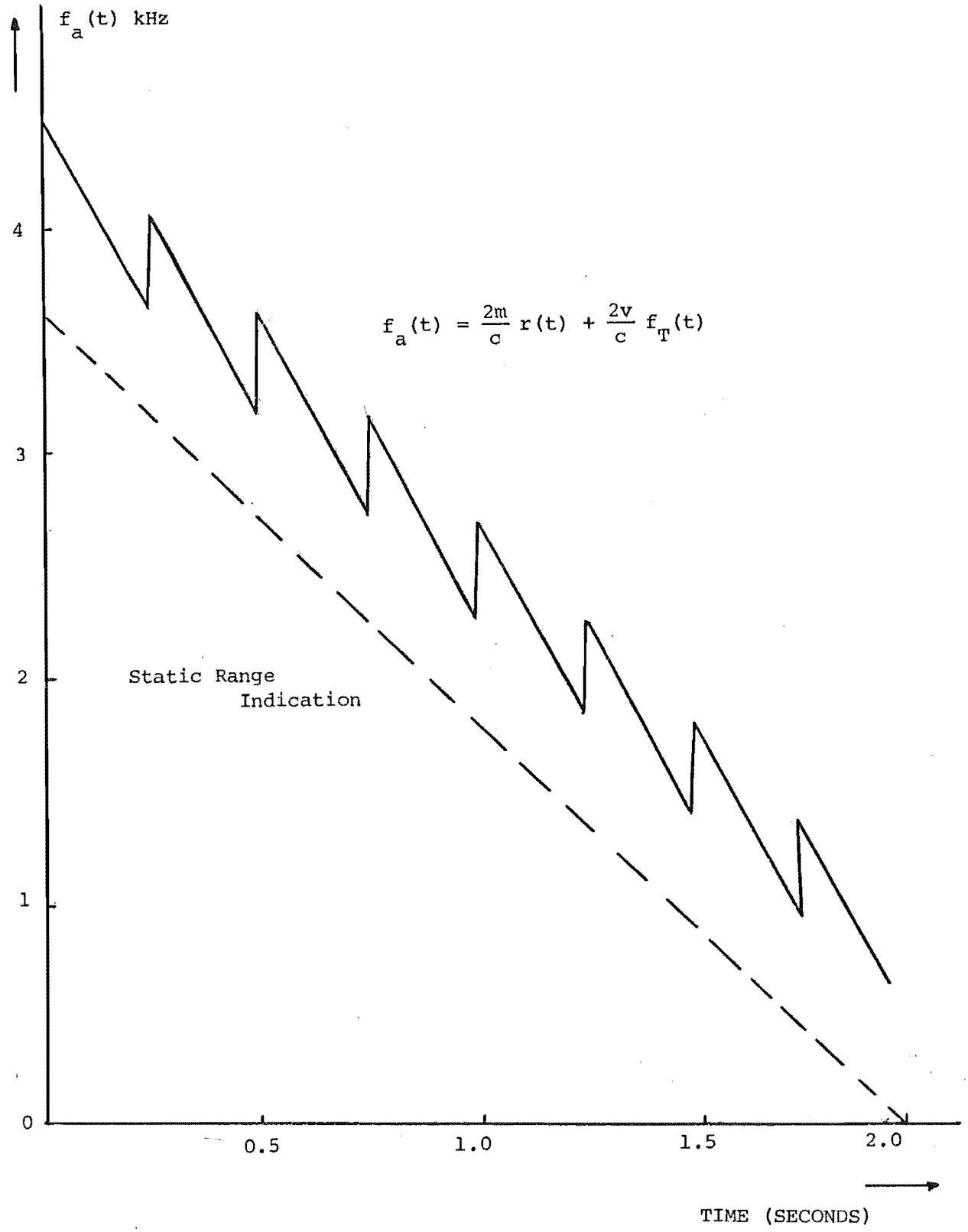


Fig. 2. Audio Frequency Variation Given by a Sensory Aid Due to Constant Radial Velocity - v = 1.5 m/s , f_2 = 100 kHz, f_1 = 50 kHz, m = 200 kHz/s, T_s = 250 ms . The dashed line corresponds to the static range indication)i.e. no Doppler effect).

with respect to the sonar system. For example the herring may migrate at a speed of up to 75 cm/s [9]; the sockeye salmon migrate at about from 20 cm/s to 53 cm/s [10,11]. While tracking a shoal, the boat does not travel faster than 3 m/s (six knots), hence a variation in the relative range rate of about 30% is very likely.

To describe the audible sound pattern produced by the CTFM sonar in a varying velocity situation, the general expression of the delay time $\tau(t)$ must be used:

$$\frac{\tau(t)}{2} = \frac{r(t_r)}{c} \tag{15}$$

where t_r is the instant when the signal transmitted $\tau(t)$ seconds ago hits the target.

$$t_r = t - \frac{\tau(t)}{2} \tag{16}$$

or

$$\frac{\tau(t)}{2} = \frac{1}{c} r(t - \frac{\tau(t)}{2}) \tag{17}$$

the right hand side of equation (17) is a function of an unknown function, $\tau(t)$, so it can not be solved directly. However if the range rate $\dot{r}(t)$ is much smaller than the sound velocity c, as discussed in section 3.2.1, then the approximation

$$\frac{\tau(t)}{2} \simeq \frac{r(t)}{c} \tag{18}$$

meets an error of only

$$\frac{\Delta\tau(t)}{\tau(t)} = \frac{1}{c} \frac{r(t) - r(t - \frac{\tau(t)}{2}}{\frac{\tau(t)}{2}} \tag{19}$$

$$\frac{\Delta\tau(t)}{\tau(t)} = \frac{\dot{r}(t)}{c} < 0.5\% \tag{20}$$

The approximation in equation (18) is reasonable, so:

$$\dot{\tau}(t) \simeq \frac{2\dot{r}(t)}{c} \tag{21}$$

Substituting these equations into equation (10), the range coding frequency will have the form:

$$f_a(t) \simeq \frac{2m}{c} r(t) - \frac{2f_T(t)}{c} \dot{r}(t) \qquad (22)$$

For the special case of constant acceleration, a towards the observer, the range coding frequency is

$$f_a(t) \simeq \frac{2m}{c} (r(0) - v_o t - \frac{a}{2} t^2) + \frac{2}{c} (f_2 - mt_s)(v_o + at) \qquad (23)$$

where v_o is the initial velocity.

From the plot of Figure 3 it will be seen that the acceleration produces a non-linear change of audio frequency during the period $t_n < t < t_{n+1}$.

A typical audible frequency pattern perceived by a blind aid user when approaching and stopping in front of an object is shown in Figure 4.

It will be evident that under real life conditions, for example, mobility by the blind in a busy pedestrian area where people are moving with varying acceleration, and where the relative velocity between user and fixed objects will also vary, the audio signals from the sensory system will change in an indescribably complex way - each signal having its own unique character.

In the case of the fishing sonar the effect of the ship's acceleration is usually trivial, but Doppler shift produced by large fish may not be. Two kinds of motion - (i) translational motion of the entire shoal, and (ii) oscillatory motion of individual fish within the shoal, were mentioned by Smith [5]. The translational velocity of the entire shoal is frequently smaller than 75 cm/s [9, 10,11] (the corresponding frequency shift is less than 60 Hz). The variation of relative velocity between sonar user and the shoal due to this motion may be recognizable but its effect is still not very significant in comparison to that due to oscillatory motion of individual fish within the shoal. Indeed, Bainbridge's results [12] showed that average fish (length 75 cm) like Salmon may reach

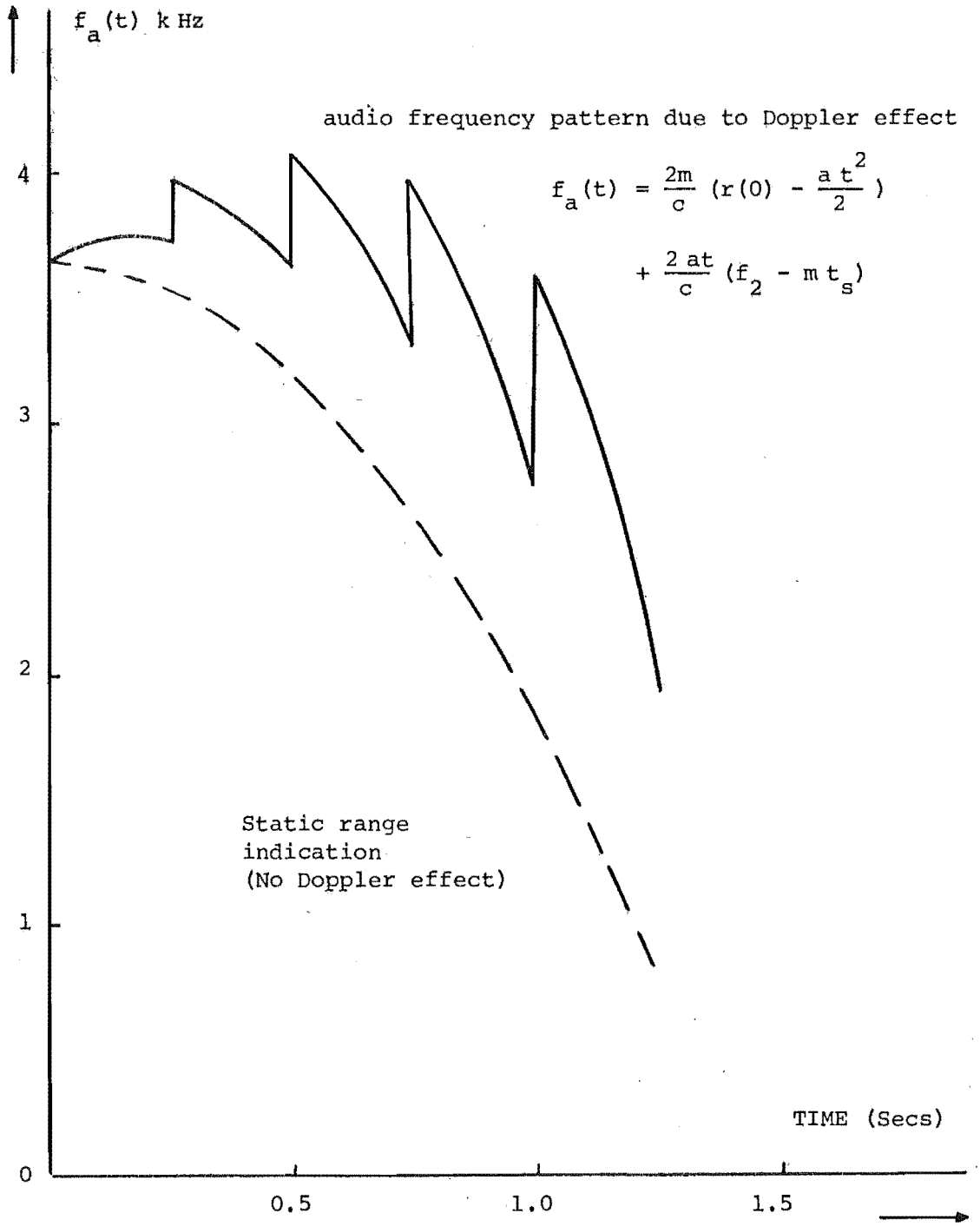


Fig. 3. Constant acceleration Audio Frequency Pattern given by the sensory aid, Initial velocity $v_o = 0$, Constant acceleration $a = 3\,m/sec^2$, $f_2 = 100$ k Hz, $f_1 = 50$ k Hz, $m = 200$ kHz/sec, $T_s = 250$ ms. The dashed line is the static range indication.

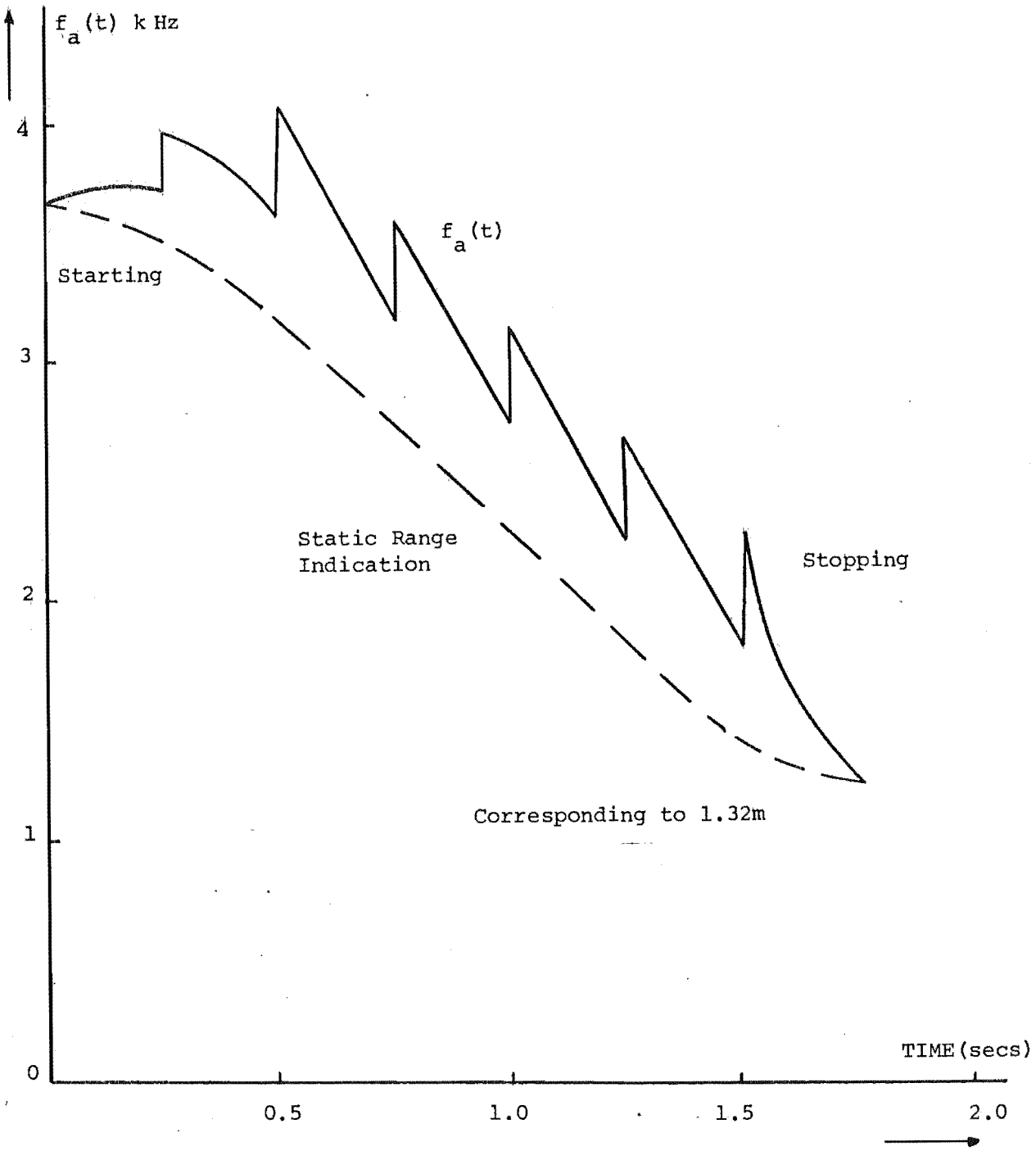


Fig. 4. Varying Velocity Audio Frequency Pattern given by the sensory aid (approaching an object from 3 m and stopping at 1.32 m). For $0 \leq t \leq 0.5_s$, initial velocity = 0, acceleration = 3 m/s^2; For $0.5_s \leq t \leq 1.5_s$, constant velocity = 1.5 m/s; For $1.5_s \leq t \leq 1.75_s$, deceleration = 6 m/s^2.

a speed of 6 m/s, and even some small fish (length 25 cm) can swim at 4 m/s though for only a short period of time. This oscillatory motion of the fish in the shoal can create the frequency shifts up to 480 Hz in the echoes if 60 kHz signals are transmitted. Using 70 kHz transmitted signal, Hester [13] observed the frequency shifts of about 200 Hz in the echoes returned from individual fish. Since these frequency shifts in the echoes are preserved in the range coding frequency of the CTFM sonar, after the echoes are demodulated, consequently an oscillatory frequency modulated sound pattern is produced. In addition to this effect amplitude modulations in the sound pattern (due to the changes in scattering cross section of fish while they are swimming) are also possibly heard.

Fish of different sizes or different kind have different types of motions. Hence the sound patterns they produce must be different. Fishermen frequently know the kinds of fish possibly found in the waters they are fishing. Also they are interested only in some specific fish which form shoals, hence they will be able to learn to recognize different species from the difference in character of the echo sounds these species produce [5].

3.3 BINAURAL SOUND PATTERNS

The discussion has thus far been about the range indication. Under realistic conditions, the relative direction of the object is always changing during the locomotion of the sensory system user. Hence the audible frequency pattern will flow left or right in accord with the changing spatial positions of the reflecting object relative to the user. The direction code of the binaural sensory system is designed to follow the relationship of

$$\theta_e = k \log \left(\frac{I_R}{I_L}\right) \tag{24}$$

as far as this is physically possible,

where θ_e is the estimated direction as perceived by the user,

k is the auditory localizing coefficient of the user,

I_R, I_L are the sound intensities of the right and left ears respectively.

Subjects vary their sensitivity to interaural amplitude difference, and by a suitable choice of k, the direction code can be matched to the user so that $\theta_e = \theta$, the actual direction. The I.A.D. code is obtained in practice by using a specially designed transmitter - receiver transducer arrangement [14]. The ideal overall transducer response is shown by Rowell [15] to be

$$A_i(\theta) = C \exp\left(- \frac{(\theta + (-1)^i \alpha)^2}{4 \alpha k}\right) \tag{25}$$

where α is the splay angle of the receiver transducers, C a constant, the index, $i = 1,2$, denotes the right and left responses respectively (Fig. 5).

The following are some typical situations frequently encountered in realistic conditions.

3.3.1 Object Passed in One Side

This is the most typical situation such as when a blind aid user walks along the footpath and hears a lamp-post passed on his left or right side. Figure 6 shows that if the blind aid user walks at a constant velocity $\vec{V}$, the relative range between him and an object will vary at a decelerated rate, $\dot{r}(t)$, determined by

$$\dot{r}(t) = -v \cos \theta(t) \tag{26}$$

since

$$r^2(t) = x^2(t) + d^2 \tag{27}$$

and

$$\theta(t) = \cos^{-1} \frac{x(t)}{r(t)} \tag{28}$$

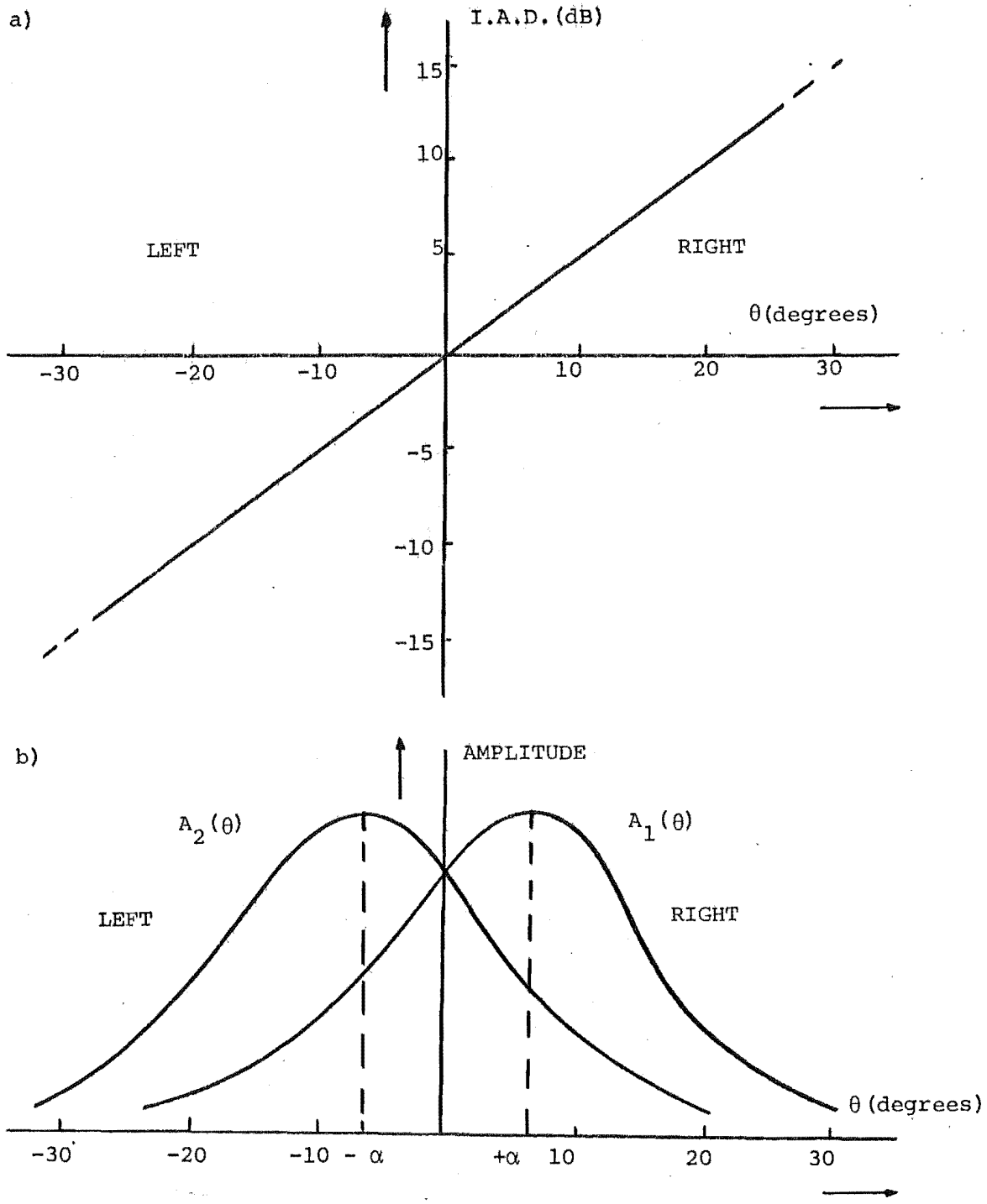


Fig. 5. - (a) I.A.D. for Idealized Audio Output

(b) Idealized Variation of Amplitude in two receivers of a binaural sonar.

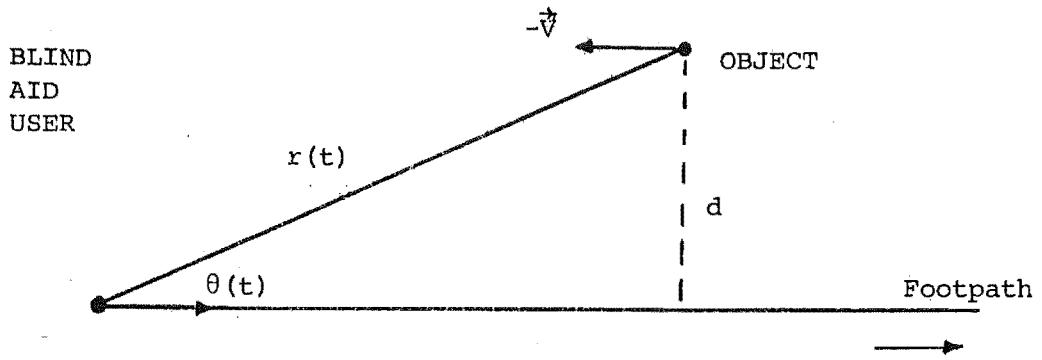


Fig. 6. Object is passed on the left side of the blind aid user.

where d is the distance of the object from the footpath,

$x(t) = x(o) - vt$ is the abscissa of the object, $v = |\vec{v}|$,

and $\theta(t)$ is the azimuthal direction of the object.

Substituting equations (26), (27) in equation (22), the binaural sound pattern can be described by

$$f_a(t) = \frac{2m}{c} r(t) + \frac{2 v \cos \theta(t)}{c} f_T(t) \qquad (29)$$

and $$I.A.D.(t) = K \theta(t) \qquad (30)$$

where K is a constant.

Figure 7 illustrates the difference in character of the sound patterns given by a central object and a side-edged object. The effect of deceleration is shown in the sound pattern of the latter.

3.3.2 The Effect of Head Rotation on the Sound Patterns

Observation of the performance of the sensory aid user approaching and walking between two rows of poles (recorded on film) reveals that the blind user frequently rotates his head to find the

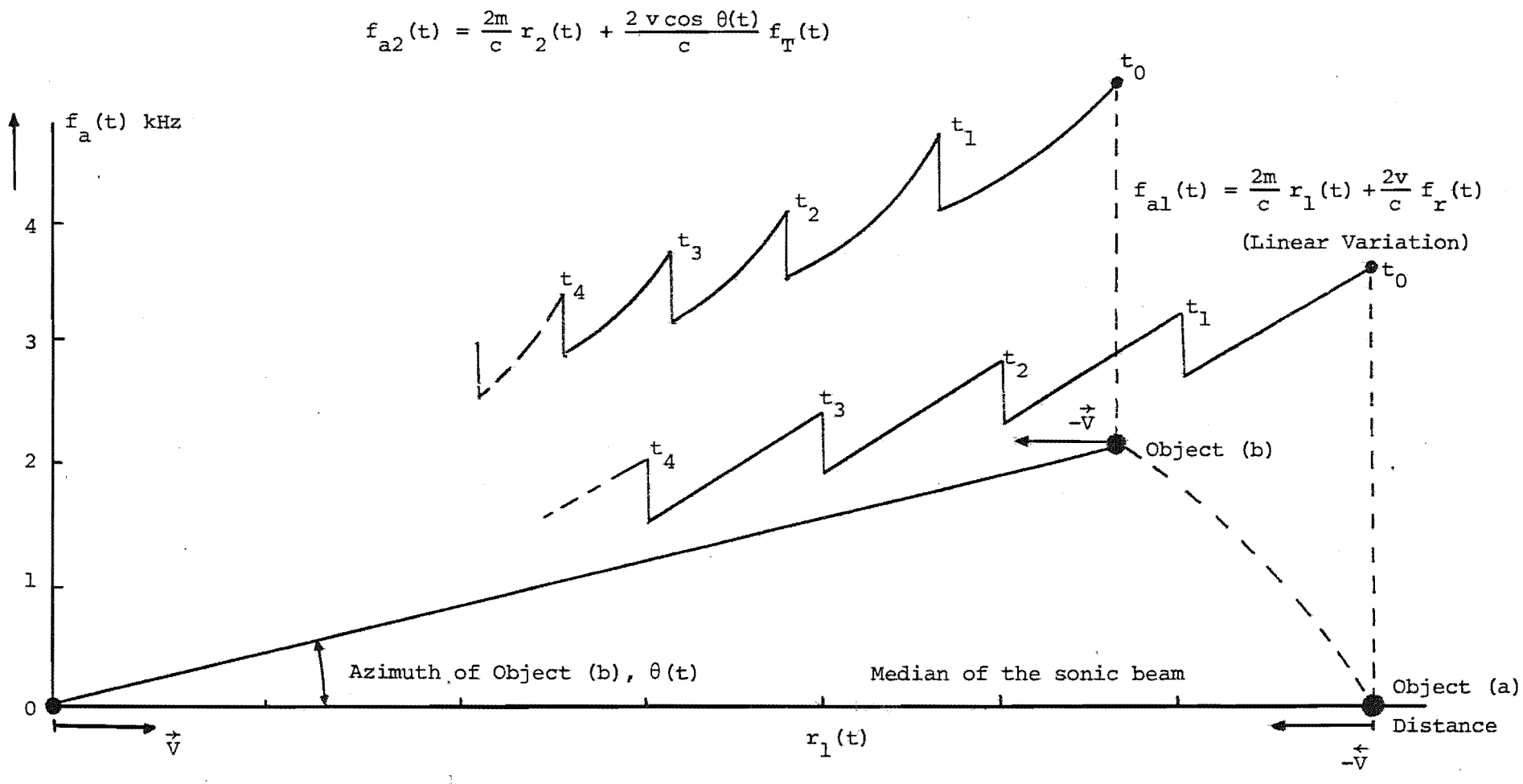


Fig. 7. Comparison between the sound patterns of a centre object and of a side object.

correct path. This kind of movement as mentioned by Rowell [15] possibly affects the resolution capability of the user. An investigation of the effect of head rotation on the sound pattern is given in this sub-section.

When the sensory aid user rotates his head, the sonar, as illustrated in Figure 8, moves laterally as well as rotates. Therefore both range and direction of object relative to the sonar vary. The instantaneous range and azimuthal direction, $r(t)$ and $\theta(t)$, of an object can be determined as follows.

Let r_b and θ_b be the initial range and azimuth of the object when the sensory system is at the position B. θ_b is positive if the object is at the right hand of the median of the sonic beam. At the time t_b, the head starts to rotate, the instantaneous rotation angle is denoted by $\alpha(t)$ which is positive when the head rotates to the left hand. a is the distance from the rotation centre, O, of the head to the sonar. From Figure 8, when the sonar is at the position P, we have

$$\theta(t) \simeq \theta_b + \alpha(t) \tag{31}$$

and

$$r(t) \simeq r_b + |\overline{BP}| \sin\left(\theta_b + \frac{\alpha(t)}{2}\right) \tag{32}$$

Provided that $r_b >> a$

Since $|\overline{BP}| = 2a \sin\frac{\alpha(t)}{2}$

$$r(t) \simeq r_b + 2a \sin\frac{\alpha(t)}{2} \sin\left(\theta_b + \frac{\alpha(t)}{2}\right) \tag{33}$$

or

$$r(t) \simeq r_b + a[\cos\theta_b - \cos(\theta_b + \alpha(t))] \tag{34}$$

The range rate is

$$\dot{r}(t) \simeq a\,\dot{\alpha}(t) \sin(\theta_b + \alpha(t)) \tag{35}$$

where $\dot{\alpha}(t)$ is the rate of rotation.

Let (t_b, t_e) be the interval of time during which the blind

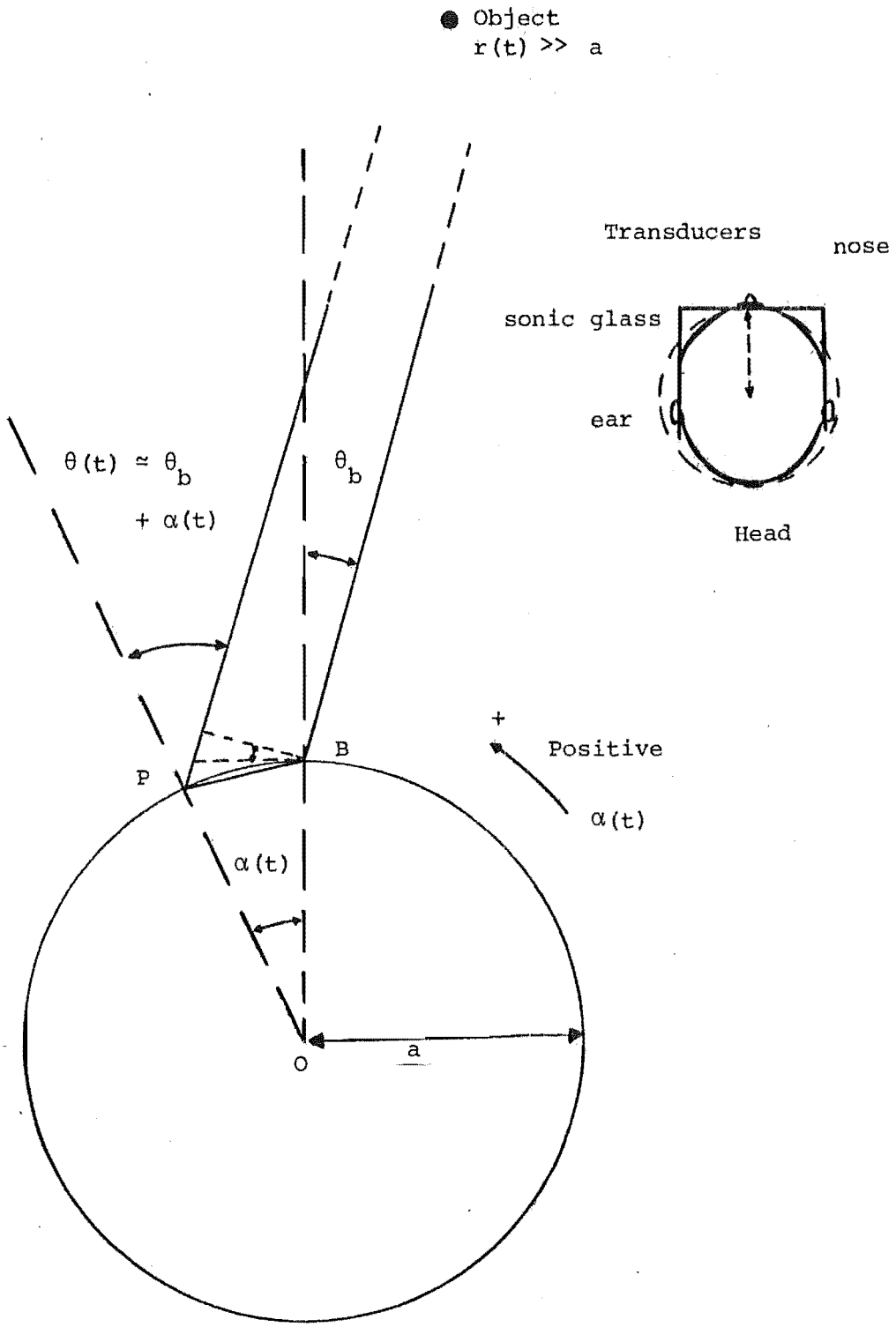


Fig. 8. Variation of Range and Azimuth of an object due to head rotation. $\theta(t)$ is positive if target is in the right of the beam axis.

user scans his head an angle of α_{max} from the forward looking position B . $\dot{\alpha}(t)$ can be approximated by

$$\dot{\alpha}(t) \simeq \frac{\alpha_{max}}{t_e - t_b} \tag{36}$$

The audible frequency pattern is obtained by substitution of equations (34) and (35) in equation (22).

$$\begin{aligned} f_a(t) &= \frac{2m}{c} r_b && t < t_b \\ f_a(t) &= \frac{2m}{c} r(t) - \frac{2f_T(t)}{c} a\,\dot{\alpha}(t)\sin(\theta_b + \alpha(t)) && t_b \leq t \leq t_e \\ f_a(t) &= \frac{2m}{c} r(t_e) && t > t_e \end{aligned} \tag{37}$$

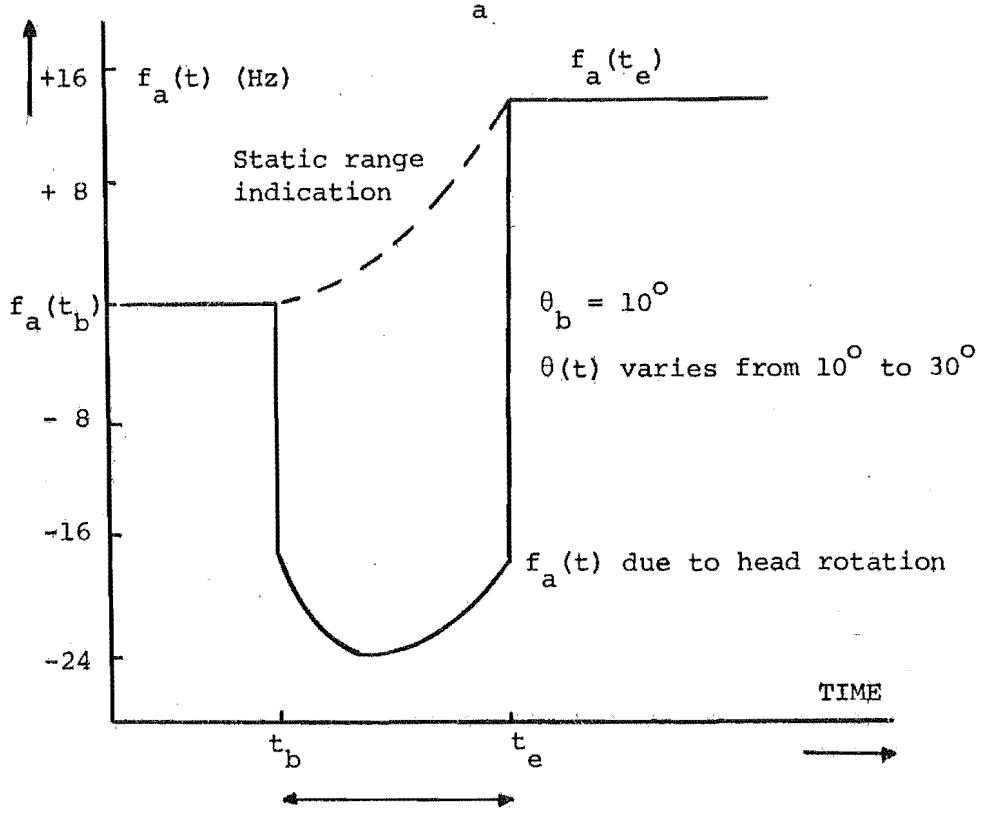
Figure 9a, b, c, d shows the audible frequency patterns produced by the sensory aid when the user rotates his head an angle of 20^{o} within 0.2 secs. The radius a is given to be 13 cm in this example (measured and averaged over several subjects). The initial azimuth of the object is given on each plot. The Doppler effect creates a negative frequency pulse when blind aid user turns away from the object, and a positive frequency pulse when he turns towards it.

In the above example, the rate of rotation is assumed to be constant. If acceleration at the beginning and decleration at the ending of the head rotation are taken into account, the frequency-pulses are less sharp than those in Figure 9.

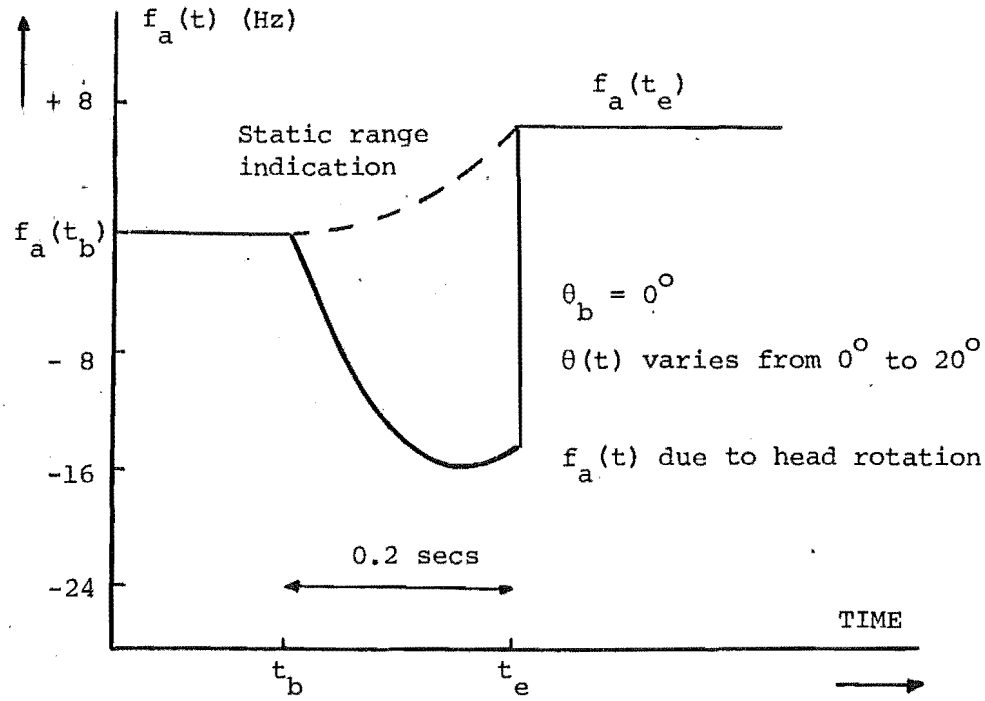
Figure 10 illustrates the possibility of discriminating two different sound patterns given by two objects different in direction but close in range, by rotating the head.

In Figure 10, the sound patterns are described by their time varying frequencies and I.A.D.'s. In fact, the capability of discrimination may also be affected by the variation in loudness

a


+16
$f_a(t)$ (Hz)
$f_a(t_e)$
Static range indication
+ 8
$f_a(t_b)$
$\theta_b = 10^o$
$\theta(t)$ varies from 10^o to 30^o
- 8
-16
$f_a(t)$ due to head rotation
-24
TIME
t_b
t_e

b


$f_a(t)$ (Hz)
+ 8
$f_a(t_e)$
Static range indication
$f_a(t_b)$
$\theta_b = 0^o$
$\theta(t)$ varies from 0^o to 20^o
- 8
-16
$f_a(t)$ due to head rotation
0.2 secs
-24
TIME
t_b
t_e

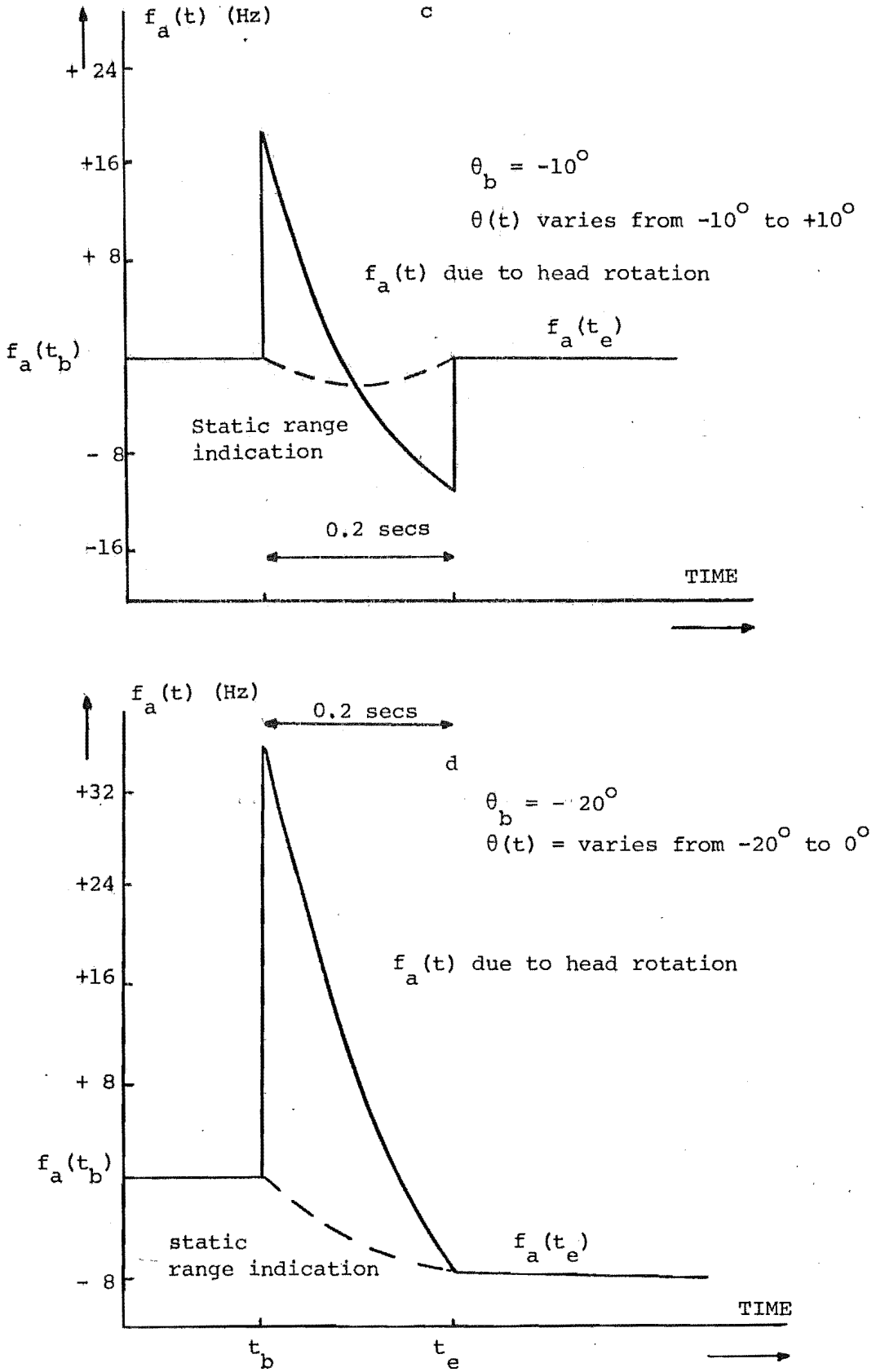


Fig. 9. The effect of head rotation on the range coding frequency. (On the axis of ordinated, $f_a(t_b)$ is the reference, so +8 means $f_a(t) = f_a(t_b) + 8$Hz.)

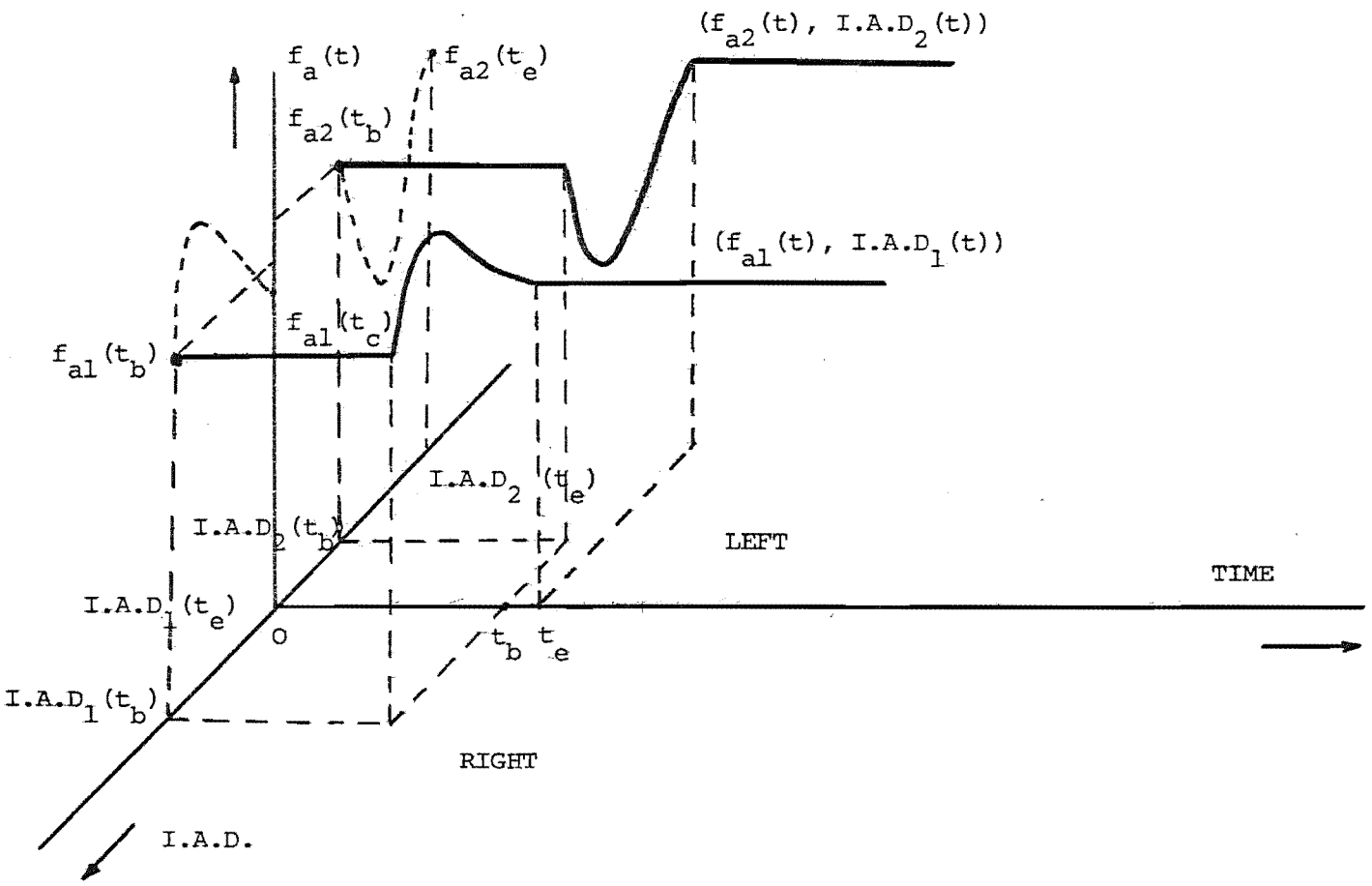


Fig. 10. The effect of head rotation on the binaural sound patterns of two objects initially close in range but different in azimuthal angle (head rotates to the right in this figure).

of the sound patterns. The transmitter beam is frequently of a bell-shape which drops 3dB at, say, $\pm 30^{\circ}$, and decreases much faster at larger azimuthal angle. Hence the head rotation of the sensory aid user may reduce the loudness of a time varying tone significantly, say, 10dB, if its corresponding object is moved to an azimuthal angle larger than 30°. In Chapter 2, it was shown that when two pure tones are simultaneously presented, the capability of resolution can be increased by reducing the loudness of the low frequency component. Conversely, if the low frequency component is of a higher level, the high tone will be masked significantly [16]. Consequently when the sensory aid user turns towards the far object, the possibility that he hears two tones increases, but when he turns towards the near object, the high tone given by the far object may be entirely masked.

In the case of the fishing sonar, the lateral shift of either the boat or the objects may create similar sound patterns.

3.4 DISCUSSION

It is experienced that under stationary conditions when two signals of nearly the same frequency are being received, one from object (a) and the other from object (b) separated by an angle $\Delta\theta$, the sensory aid user is unable to discriminate the L and R stimulus from object (a), and the L and R stimulus from object (b). Only when the L and R responses from object (a) fuse to indicate an object in direction θ_a, and the L and R responses from object (b) fuse to indicate an object in direction θ_b, is it possible to expect discrimination in direction. It is also well known that two tones presented to a subject simultaneously must differ considerably in frequency to be heard as two distinctly separate notes (Chapter 2).

It would seem therefore that spatial resolution in the binaural system would be exceedingly poor.

It has been shown however that locomotion, when wearing the sensory aid, produces rapidly varying auditory signals - not tones. Only rarely are we interested in the system resolution when standing perfectly still. The same applies to the fishing sonar - if less so. Experience with blind people using the sensory aid in real life situations suggests that much greater resolution is being enjoyed. What the blind people hear are the time varying "tones" producing flow patterns of sound related to the changing spatial positions of the reflecting objects relative to the user. Figures 7 and 10 illustrate the discrimination of sound patterns given by two objects close in range but different in direction. The former is achieved during normal locomotion while the latter is obtained by the head rotation.

The flow patterns obtained when, for example, objects are being approached and passed on the left or right are not unlike the flow patterns experienced by a driver of an automobile passing street lights in thick fog. In the case of either the sensory aid user or the driver in thick fog, the stimulus is a continuous family of transformations which is unique to the particular path of locomotion. The forms of continuous transformation are quite specific, and an experienced subject can probably discriminate among them. Gibson [17] discusses such flow patterns in vision during locomotion. Hunt [18] demonstrated that a fixed pattern of black and white, such as block printing, can be recognized visually through a moving overlay which consists of a random pattern of opacity and transparency, even when the moving pattern is dense enough to mask the signal pattern completely when the overlay is not in motion. He suggested that a similar effect

may occur in audition that a time invariant pattern of frequency response can be discriminated in the presence of a variable pattern of frequency response when the listening exposure is of sufficient duration. This suggestion may apply to a more general situation where several time varying sound patterns, each having unique character in both binaural difference and pitch quality, are simultaneously presented and possibly discriminated by an experienced subject. Thus, even though the form of the time-varying "sound patterns" produced by the sensory system may be mathematically very complicated for realistic mobility situations when several objects are in the field of view at once, each pattern must be distinctly perceived as having a very simple cognitive form related to the changing spatial coordinates.

To describe quantitatively the resolution capability of the sensory system as a whole, a series of controlled experimental tests have been conducted simulating dynamic conditions. These are reported in the next chapter.

REFERENCES

[1] KAY, L. A new or improved apparatus for furnishing information as to the position of objects. Patent Specification No. 978741. The Patent Office, London, 1959.

[2] KAY, L. Blind Aid. Patent Specification No. 3366922. United States Patent Office, Washington, D.C., 1965.

[3] MARTIN, G. Electronics and transducers for an ultrasonic blind mobility aid. M.E. Thesis, University of Canterbury, 1969.

[4] SMITH, R.P. and KAY, L. A fish-finding sonar utilizing an audio information display. Digest of Technical Papers, IEEE Ocean Conf., Panama City, Florida, 1970, p.113.

[5] SMITH, R.P. Transduction and audible displays for broad band sonar systems. Ph.D. Thesis, University of Canterbury, 1973.

[6] KAY, L. A sonar aid to enhance spatial perception of the blind: Engineering design and evaluation. The Radio and Electronic Engineer, Vol.44, 1974, p.605.

[7] AIRASIAN, P. Evaluation of the binaural sensory aid for the blind. A.F.B. Research Bulletin, Vol.26, 1973, p.51.

[8] KAY, L. Sonic glasses for the blind - presentation of evaluation data. A.F.B. Research Bulletin, Vol.26, 1973, p.35.

[9] HARDEN JONES, F.R. Fish migration. London, Edward Arnold Publishers Ltd., 1967, pp.230-234.

[10] MADISON, D.M. et al. Migratory movements of adult Sockeye salmon (oncorhyn chus verka) in coastal British Columbia as revealed by ultrasonic tracking. J.Fish.Res.B.Can., Vol.29, 1972, pp.1025-1033.

[11] GROOT, C. Migration of yearling Sockeye salmon (oncorhyn chus verka) as determined by time-lapse photography of sonar observations. J. Fish. Res. B. Can., Vol.29, 1972, pp.1431-1444.

[12] BAINBRIDGE, R. The speed of swimming of fish as related to size and to the frequency and amplitude of the tail beat. J. Exp. Biol., Vol.35, 1958, pp.109-133.

[13] HESTER, F.J. Identification of biological sonar targets from body-motion doppler shifts. Symposium on Marine Bio-Acoustics, New York, Pergamon Press, Vol.2, 1967, pp.59-74.

[14] KAY, L. Toward objective mobility evaluation, some thoughts on a theory. A.F.B., New York, Monograph, 1974.

[15] ROWELL, D. Auditory display of spatial information. Ph.D. Thesis, University of Canterbury, 1970.

[16] WEGEL, R.L. and LANE, C.F. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys.Rev., Vol.23, 1924, pp.266-285.

[17] GIBSON, J.J. Visual controlled locomotion and visual orientation in animals. British Journal of Psychology, Vol.49, 1958, pp.182-194.

[18] HUNT, F.V. Role of time integration in the discrimination of frequency response patterns. J. Acoust. Soc. Am., Vol.34, 1962, p.125.

[19] KAY, L. and DO, M.A. An artificially generated multiple object auditory space for use where vision is impaired. Acustica, Vol.36, 1976/77, pp.1-8.

CHAPTER 4

A DEFINITION OF FREQUENCY RESOLUTION USING NEW AUDITORY SENSATIONS

CHAPTER 4

A DEFINITION OF FREQUENCY RESOLUTION USING NEW AUDITORY SENSATIONS

4.1 INTRODUCTION

In the previously published paper by Kay and Do [1], and also in Chapter 3, it was shown that auditory signals produced by a CTFM wide beam, very wide-band sonar vary in a manner which is uniquely related to spatial change. It was suggested that the patterns of change which are produced by relative movement could be the reason why blind users of this sensory system are able to discriminate the complex signals from multiple objects in a real environment. This was deduced from the observed ability of blind people to negotiate complex real environments and be more effectively mobile [2]. No supporting measures of a quantitative nature were obtained during the extensive evaluation [3] because of the difficulty in collecting meaningful data on what subjects were using as spatial cues under dynamic conditions. It is, of course, possible to measure performance in a simple well controlled situation [4] but once an individual is allowed free movement and personal control of his motion in his normal habitat, the variables become impossible to handle.

Laboratory experiments prior to the evaluation, which were designed using system simulation to determine the ability of subjects to resolve objects in space by auditory means, used only static situations to facilitate controlled measurements [5]. The results of these experiments showed subjects had poor resolution capability (see Chapter 2) and failed to explain subjective impressions of resolution capability whilst in motion when using the sensory system. Rowell [5]

did, however, show that the simulated direction cue of interaural amplitude difference could be matched to individuals more effectively when rotational head movement was permitted to vary the cue. Subjects could estimate direction more accurately when relatively angular motion took place. It is pointed out in Chapter 3, the possibility of discriminating the sound patterns from two objects close in range but adequately separated in direction.

In the case of the fishing sonar designed by Smith [6] the ability of an operator to discriminate between two shoals of fish appearing simultaneously in the field of view remained unexplained. Here the operator is presented with reverberation from all ranges simultaneously, the "cluster" of tones - one from each fish - in shoal (a) in the mean direction θ_a and the "cluster" of tones from shoal (b) in the mean direction θ_b. He was able to say which shoal was the larger and which was the nearer of the two. Experiments to determine auditory resolution under these conditions will require the generation of many tones having slightly differing spatial characteristics which include the random spatial motion of fish within the shoal. True production of such a set of sophisticated conditions has not been fully given yet but many approaches which have recently been attempted will be presented in the later chapters.

This chapter describes some experiments using simulation techniques as a first attempt to determine the apparent influence of spatial change on resolution in an artificially generated auditory space. Two simple objects are simulated and the relationship to the sensory aid for the blind and the fishing sonar are discussed. A new auditory sensation is found and employed in defining the auditory frequency resolution highly performed by subjects in comparison with their performances in static cases. It will be shown that the Doppler

effect, involved in doubling the rate of change of the audio frequency, plays a significant part in this improvement. An increase in relative velocity of an object may also enhance the resolution capability quantitatively as well as qualitatively to an extent that signals may be resolved cognitively before the subject's reaction can take place. Most of the parts of this chapter have been published elsewhere [7].

4.2 RESOLVING TWO CHANGING TONES BY USING NEW AUDITORY SENSATIONS

Consider now the simulation of two adjacent objects, 1 and 2, in auditory space described by two pairs of variables, $(r_1(t), \theta_1(t))$ and $(r_2(t), \theta_2(t))$, where $r(t)$ and $\theta(t)$ are the instantaneous object range and azimuth angle respectively. Motion of an object with respect to the sonar system (Binaural Sensory Aid or Fish Sonar) changes the frequency of the transmitted wave as it is reflected by a factor of approximately $(1 - \frac{2\dot{r}(t)}{c})$ where $\dot{r}(t)$ is the relative velocity of the object and c is the velocity of propagation of the sound wave. Then for a broad band transmission signal higher frequencies must undergo larger Doppler shifts than the low frequencies of the transmission band. The range coding of the sonar system for an object of range $r(t)$ is given by

$$f_a(t) \simeq \frac{2m\, r(t)}{c} - 2\, f_T(t)\, \frac{\dot{r}(t)}{c} \qquad (1)$$

where $\dot{r}(t)$ is the time derivative of $r(t)$.

(See Chapter 3, and Kay and Do [1])

The interaural amplitude difference IAD(t) is also a function of time under conditions of motion and is proportional to $\theta(t)$. Thus when these spatial codes are used as auditory stimuli the resolution

capability must be a function of four variables [$f_{a1}(t)$, $IAD_1(t)$, $f_{a2}(t)$, $IAD_2(t)$]. Clearly any function of four dynamic variables is too complicated to be studied and described by subjects at this stage. Hence in the experiment to be described f_{a2}, IAD_2 and IAD_1 are held constant, only f_{a1} is varied as in equation 1. The experimental procedure was based entirely upon the "psychological phenomenon" described below.

The Psychological Meaning of Auditory Resolution Under Conditions of Change

One of the difficulties in psychophysical experiments is to determine a suitable method of verifying the psychological judgements of subjects to ensure that they respond to the psychophysical phenomena to be studied. It was known [5] that presenting simultaneously two tones assigned values f_{a1} and f_{a2} each with its own value of IAD produces a result of frequency resolution similar to that obtained in the experiment described in Chapter 2, which uses two special tests to ensure that (i) subject perceives only two tones, and (ii) the perceived tones are the stimuli. The assignment of a value of IAD by Rowell [5] did not affect the capability of resolution in range under static conditions as compared with zero IAD as used in the experiment of Chapter 2, but it did more clearly define the psychophysical phenomena of frequency resolution, since subjects were required to indicate LEFT or RIGHT directions to the second tone. The two tones had to be perceived separately to do this.

This same decision was required in the dynamic experiment. It was designed to simulate conditions experienced when using the sonars. Initially at $t = 0$, $f_{a1}(t) \simeq f_{a2}$ and the subject hears slow beats and feels only one complex image in the auditory space. Its IAD is the combination of IAD_1 and IAD_2. As t increases $f_{a1}(t)$

decreases according to equation (1) shown at the beginning of this section. The interference between two tones changes in a unique way (because of the signal uniqueness) until a certain instant $t = t_1$ the subject feels the image of $f_{a1}(t)$ flow "Left" or "Right" in auditory space, according to the assigned values of IAD_1 and IAD_2. The attention appears to be captured by the "moving" tone which may be thought of as forming a flow pattern in space. At this instant the subject is required to respond by pressing a switch to store the value of $f_{a1}(t)$. $f_{a1}(t)$ is allowed to continue when the attention can be transferred to f_{a2} which then appears in its appropriate auditory spatial position. This is always secondary to the initially perceived flow of $f_{a1}(t)$, but occurs almost simultaneously. The psychological phenomenon is described in Fig. 1. The judgement of subjects can be verified by questioning them on the relative movement of the images. The frequency resolution can be defined here as the frequency difference between f_{a2} and $f_{a1}(t)$ at which the two tones separate and "slide" in auditory space to their respective positions.

4.3 SIMULATION OF THE PROBLEM

We are concerned here with simulating a situation which relates to real experience. In the experimental situation described, the auditory sensations resemble what would be obtained if one object at range (r_1, θ_1) approached the sonar with constant radial velocity v, $r_1(t) = r_1(0) - vt$, and a second object were stationary at a range $r_2 = r_1(0)$ and in the direction θ_2. Then the range coding frequency of the moving object is given by

$$f_{a1}(t) = \frac{2m\, r_1(0)}{c} + \frac{2v}{c} f_2 - \frac{2v}{c} m\, t_s - \frac{2v}{c} m\, t \qquad (2)$$

$$0 \leq t_s \leq T_s$$

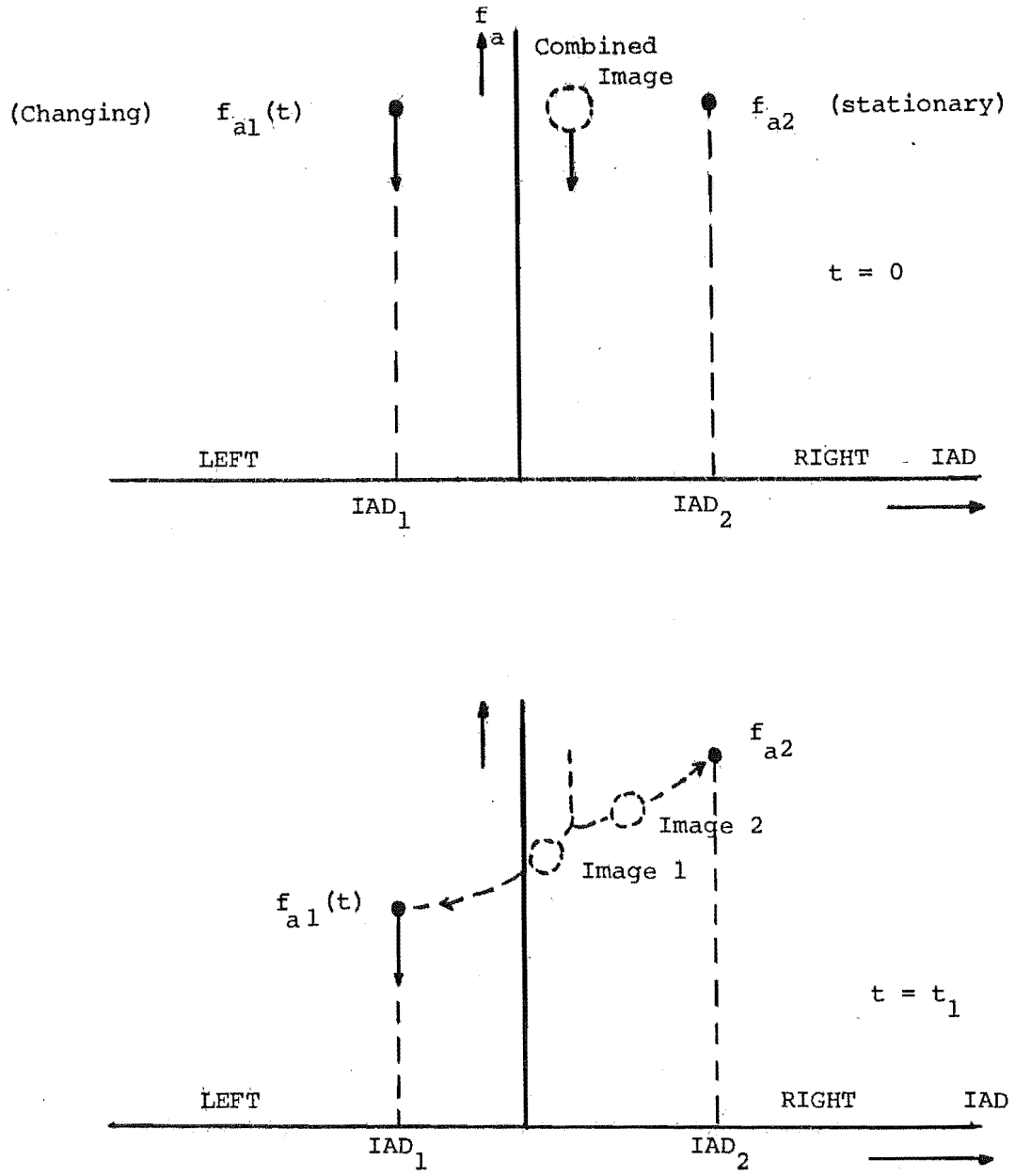


Fig. 1 The Psychological Phenomenon of Resolution in a Changing Situation

or $$f_{a1}(t) = f_{a1}(0) = \frac{2v}{c} m\ (t + t_s) \qquad (3)$$

A linear voltage controlled oscillator was used to produce this frequency function from the voltage function

$$V_{a1}(t) = V_{a1}(0) - \alpha(t + t_s) \qquad (4)$$

where α and $V_{a1}(0)$ are linearly related to $\frac{2v_m}{c}$ and $f_{a1}(0)$. An analogue computer was used to produce the voltage function of equation (4) and to control the experiment.

Figure 2 shows the panel patching of the computer and Figure 3 shows the arrangement of the equipment for the experiment.

4.4 EXPERIMENTAL PROCEDURE AND RESULTS

The subject's task was to listen to the complex sound including $f_{a1}(t)$ and f_{a2} as described, and press the switch to store the value of $f_{a1}(t)$ at the instant the descending tone was perceived to slide to the left or the right of the initial complex image. Each subject was questioned about the direction of shift of either the low tone ($f_{a1}(t)$) or the high tone (f_{a2}). For each value of f_{a2}, the experimenter recorded the value of $f_{a1}(t)$ for different values of IAD_1 and IAD_2. These latter are changed in a random manner according to a table.

4.4.1 Experiment 1. (The simulation relates to the fish sonar because the rates of change encountered are such as to allow a subject to respond without introducing serious error due to the delay in response.)

Suppose that the approach velocity of object 1 is $v = 1.5$ m/s, the velocity of sound is taken as 1500 m/s, the repetition period $T_s = 3.2$ sec, and the sweep slope $m = 12.5$ k Hz/sec then

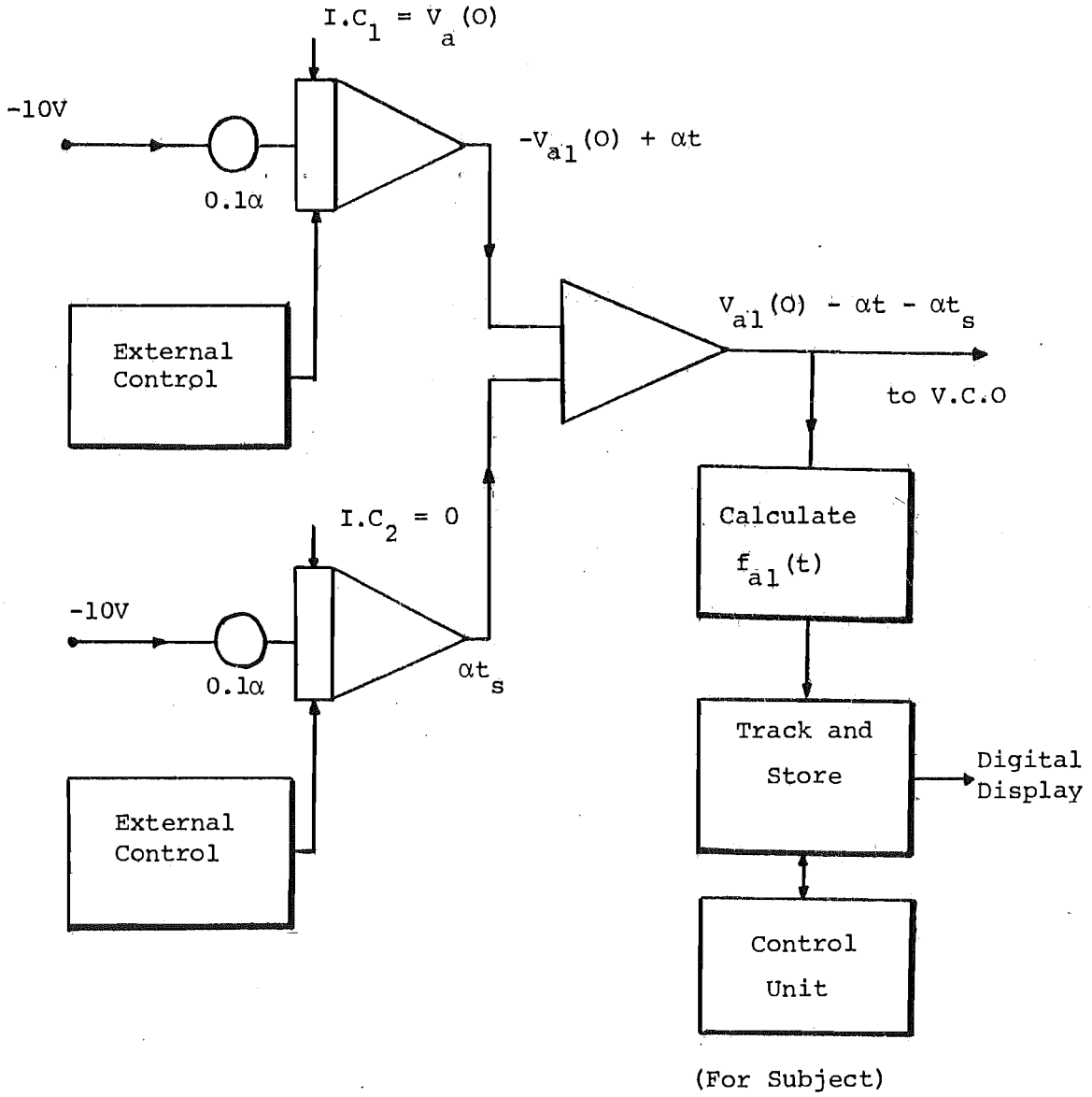


Fig. 2 - Analogue Computer to generate $V_a(t)$ and to Control the experiments

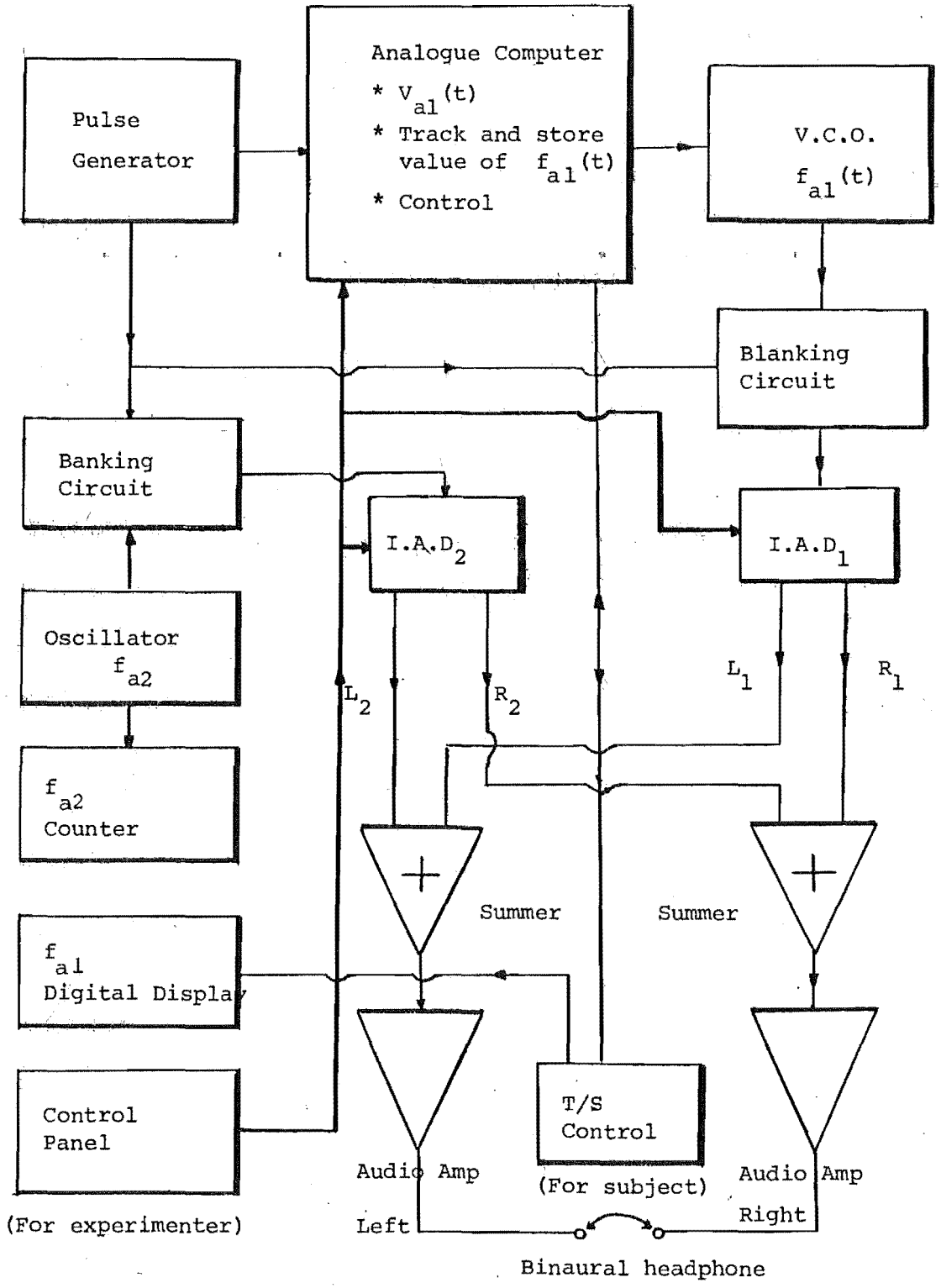


Fig. 3 - Arrangement of Equipment for the experiments

$$f_{a1}(t) = f_{a1}(0) - 25(t + t_s) \qquad (5)$$

$$0 \leq t_s \leq 3.2 \text{ s}$$

Figure 4 describes the variation of $f_{a1}(t)$ with respect to time. Three subjects S_1, S_2, S_3 were used in the experiment. Values of $f_{a1}(0) = f_{a2}$ were chosen to be 1000, 2000, and 3000 Hz, while the values of IAD were 0, +4, +8, +12 dB for object 1 and -10, -6, -2, +2, +6 and +10 dB were chosen for object 2,

In this experiment a delay in stopping $f_{a1}(t)$ of say an upper limit of 0.5 sec, due to subject's reaction time, may introduce an error of -25 Hz to the recorded value of $f_{a1}(t)$. However, since $f_{a1}(t)$ decreases in a saw-tooth manner with a sweep of 160 Hz per sweep period T_s and a reset of 80 Hz, the subject's delay may occur at the end of a sweep period when $f_{a1}(t)$ is reset 80 Hz to start a new sweep. The error may then be 80-25 = + 55 Hz. To reduce these errors $f_{a1}(t)$ was recorded 4 times and averaged. The variation of $f_{a1}(t)$ and $\Delta f = f_{a2} - f_{a1}(t)$ due to delay and reset errors and subject variation was greater than the variation between values of IAD (see Appendix 1). The dependence of frequency resolution on the IAD cannot therefore be determined. It is now assumed to be negligible so the value of Δf for each f_{a2} is found by averaging over all values IAD_1 and IAD_2. Figure 5 shows the results of the three subjects.

Quite a wide variation between subjects can be expected, but "within" subject variation is seen to be small.

4.4.2 Experiment 2. (Frequency resolution with cyclic variation removed).

The purpose of this experiment is to determine the influence of

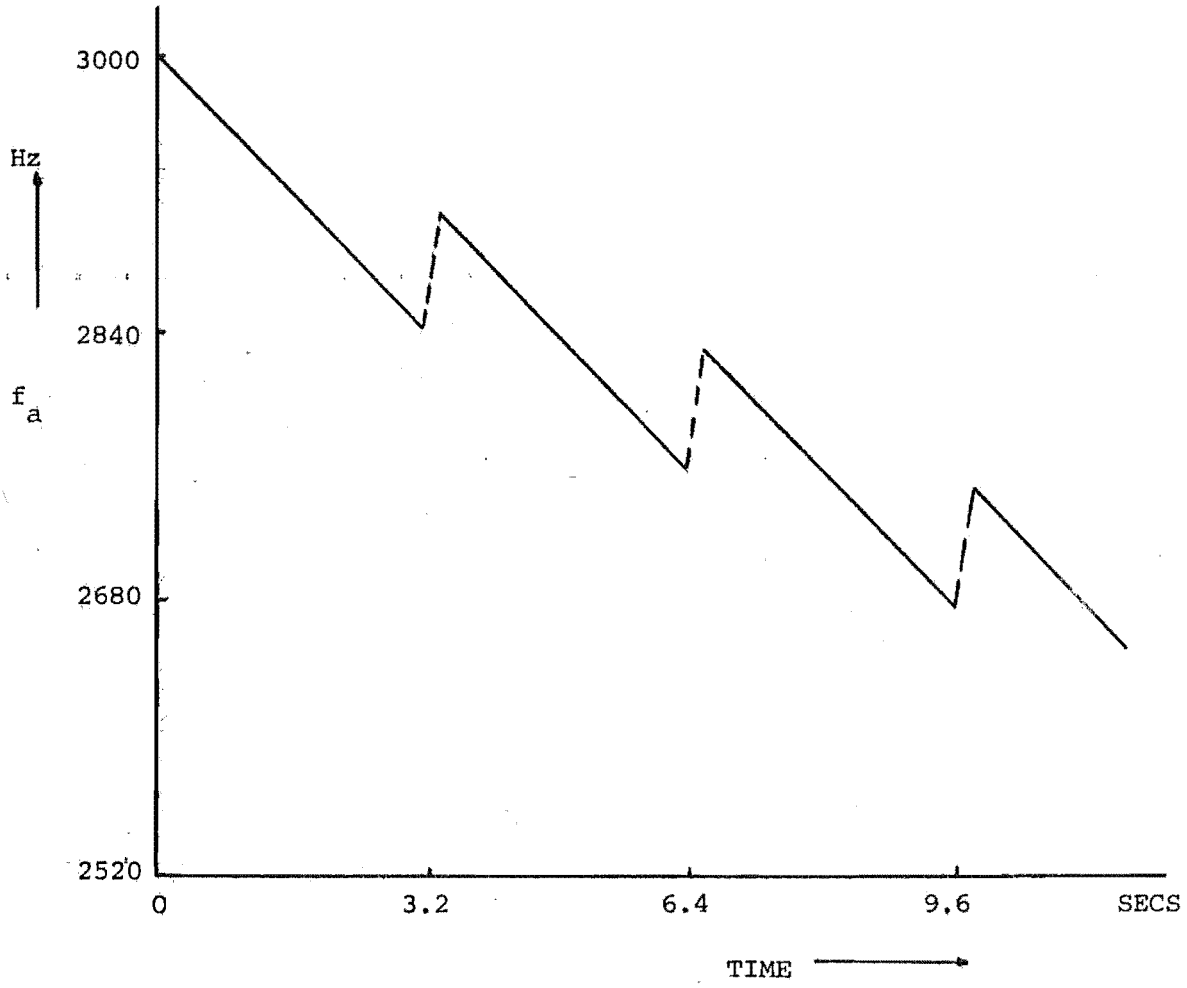


Fig. 4. Variation of Range Coding Frequency With Respect to Time in Experiment 2.

$f_{a1}(t) = 3000 - 25t - 25\,t_s$, where $0 < t_s < 3.2$. The broken lines correspond to ambiguous signals blanked out.

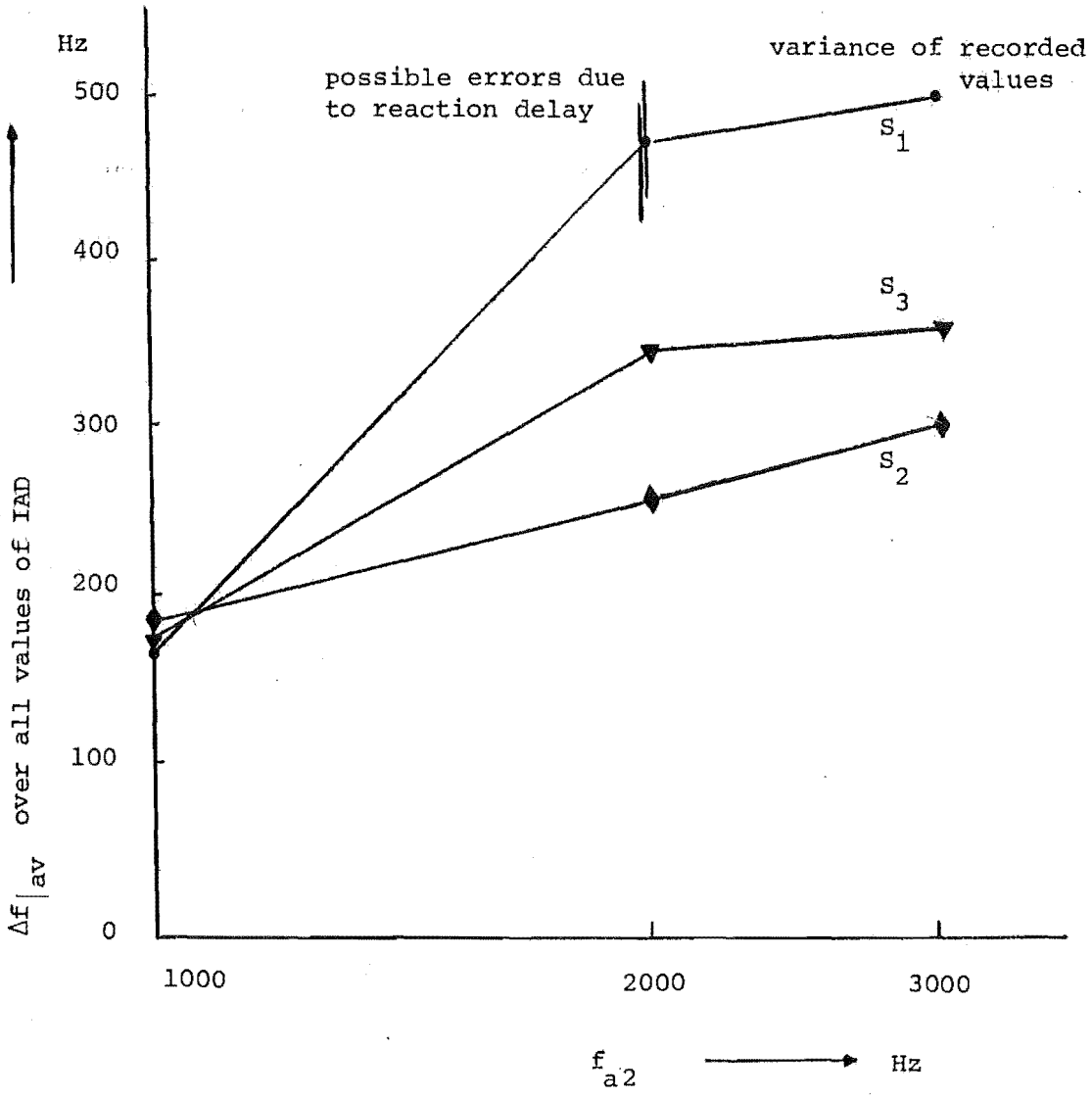


Fig. 5 - Variation of the frequency resolution with respect to frequency averaged over all values of IAD (IAD variable has negligible effect)

the saw-tooth variation on the resolution capability of the auditory system. Whilst in practise this cannot be eliminated, and it plays an important part in sharpening the ambiguity function of the sonar system, its influence on the auditory resolution could not be predicted. The frequency function for this experiment became

$$f_{a1}(t) = \frac{2m}{c} r(t) = f_{a1}(0) - \frac{2V}{c} mt \qquad (6)$$

Using the same parameters as in experiment 1

$$f_{a1}(t) = f_{a1}(0) - 25t \qquad (7)$$

Repeating the experiment with S_2, who demonstrated a higher resolution capability in experiment 1, the following result was obtained.

TABLE 1

FREQUENCY RESOLUTION WITH CYCLIC VARIATION REMOVED

f_{a2} Hz	Δf Hz	Increase in Δf %
1000	220	170
2000	440	170
3000	630	210

4.4.3 Experiment 3. (The effect of increasing the approach velocity).

Since the cyclic variation, involving doubling the rate of change of the audio frequency within each sweep, had a significant effect on the resolution capability, a direct increase in the object velocity, resulting in a higher rate of change of frequency had to have certain effects on the resolution performance of the subject. This also increased the possible error due to delay in subject's response. In this experiment, the object velocity was doubled to

3 m/s doubling the slope and the reset value as seen in Fig. 4. The following result was obtained again by S_2 .

TABLE 2

EFFECT OF INCREASING VELOCITY ON FREQUENCY RESOLUTION

f_{a2} Hz	Δf Hz	Decrease in Δf %
1000	130	30
2000	230	11
3000	240	16

Quantitatively the improvement in resolution is not significant except at low frequency. Qualitatively however, the sensation of image slide is more pronounced and the separation of images is very clear by all who experience the phenomenon. S_2 was unable to respond to change greatly the measured resolution capability obtained in experiment 1.

Table 3 summarizes all the results given by S_2 in the above three experiments and compares them with his performance recorded previously in the static case (Chapter 2).

TABLE 3

COMPARISON OF FREQUENCY RESOLUTION IN FOUR EXPERIMENTS

f_{a2} Hz	$\Delta f(0)$ Hz	$\frac{\Delta f(0)}{f}$	$\Delta f(1)$ Hz	$\frac{\Delta f(1)}{f}$	$\Delta f(2)$ Hz	$\frac{\Delta f(2)}{f}$	$\Delta f(3)$ Hz	$\frac{\Delta f(3)}{f}$
1000	400	0.50	180	0.20	220	0.25	130	0.14
2000	700	0.42	250	0.13	440	0.25	230	0.12
3000	1100	0.45	300	0.11	630	0.23	240	0.08

$\Delta f(0)$ stationary tones, f_{a1} and f_{a2}

$\Delta f(1)$ saw-tooth frequency decrease $f_{a1}(t)$

$\Delta f(2)$ Linear frequency decrease $f_{al}(t)$

$\Delta f(3)$ Saw-tooth frequency decrease $f_{al}(t)$ at double rate.

4.5 AN ATTEMPT TO IMPROVE FREQUENCY RESOLUTION BY USING AMPLITUDE-FREQUENCY WEIGHTING FUNCTION

During the above experiments, subjects frequently reported that the low-tone, $f_{al}(t)$, was heard louder when two tones were separated. This phenomenon could not be explained by the difference in loudness sensitivity of two equal intensity tones since measurements done by several experimenters had agreed that within the range of frequency below 3000 Hz , the high tone is more sensitive than the low tone [8,9, 10]. Perhaps the best explanation is the experimental results given by Wegel and Lane [11], and Fletcher [12,13] that a pure tone can mask another tone of higher frequency more easily than it can do with a lower tone. Their results also show that if two pure tones are simultaneously sounded, and the high tone is of a frequency and an intensity respectively equal to 1200 Hz and 80 dB, then a frequency difference of over 600 Hz is required so that both the above high tone and the 80 dB low-tone can be perceived without any interference-tone. But if the intensity of the low tone is reduced to 45 dB, a frequency-difference of less than 100 Hz is required for the suppression of all the interference-tones. It was this result which has aroused the idea of attempting to use an amplitude-frequency weighting function to improve frequency resolution.

Experiment 4 (Using amplitude-frequency weighting function)

This experiment is basically similar to the experiment 1. The only alteration is to feed the output of the V.C.I in Figure 3 to an amplitude-frequency weighting to reduce the intensity of the $f_{al}(t)$

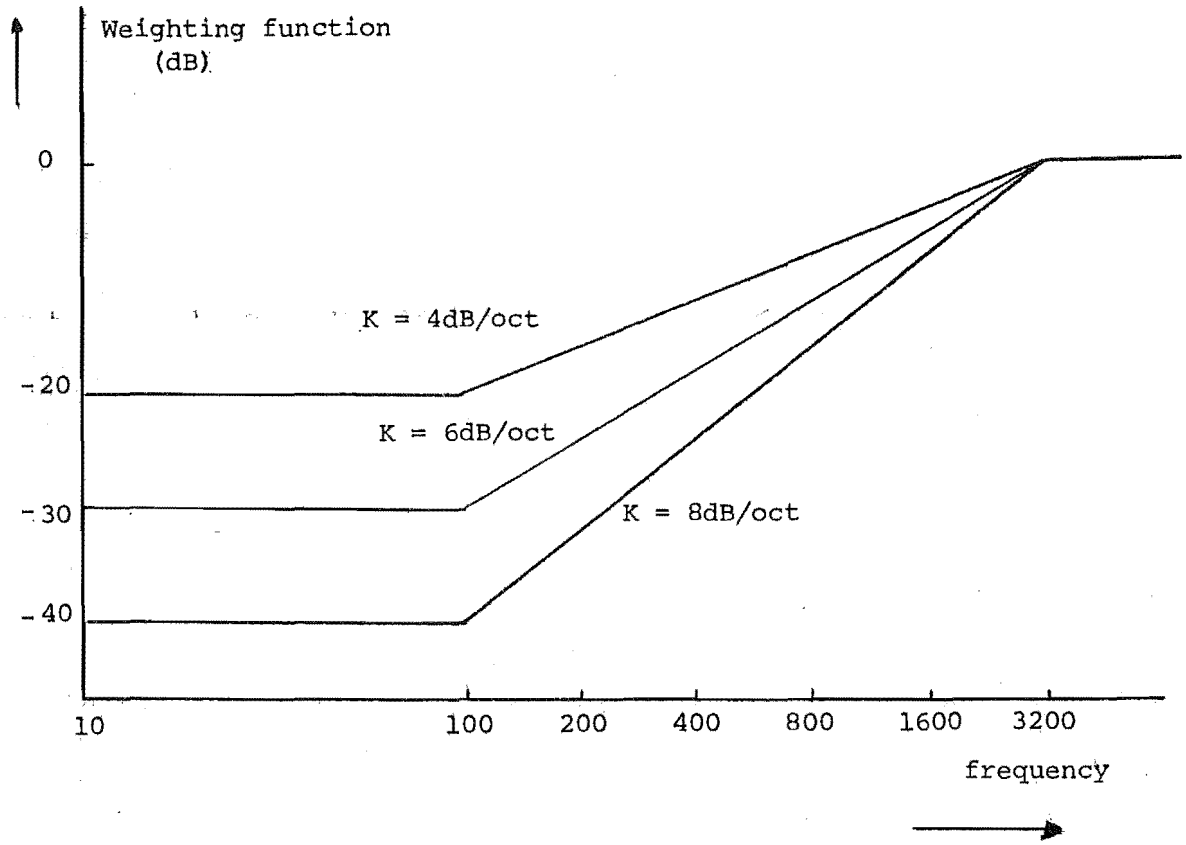


Fig. 6 - Weighting functions used in experiment 4. On linear scale the amplitude frequency variation follows the function

$$y = C\ 10^{\frac{K}{\log 2}\ \log f}$$

where $20 \log C = -\frac{K}{\log 2}\ \log\ 3200$

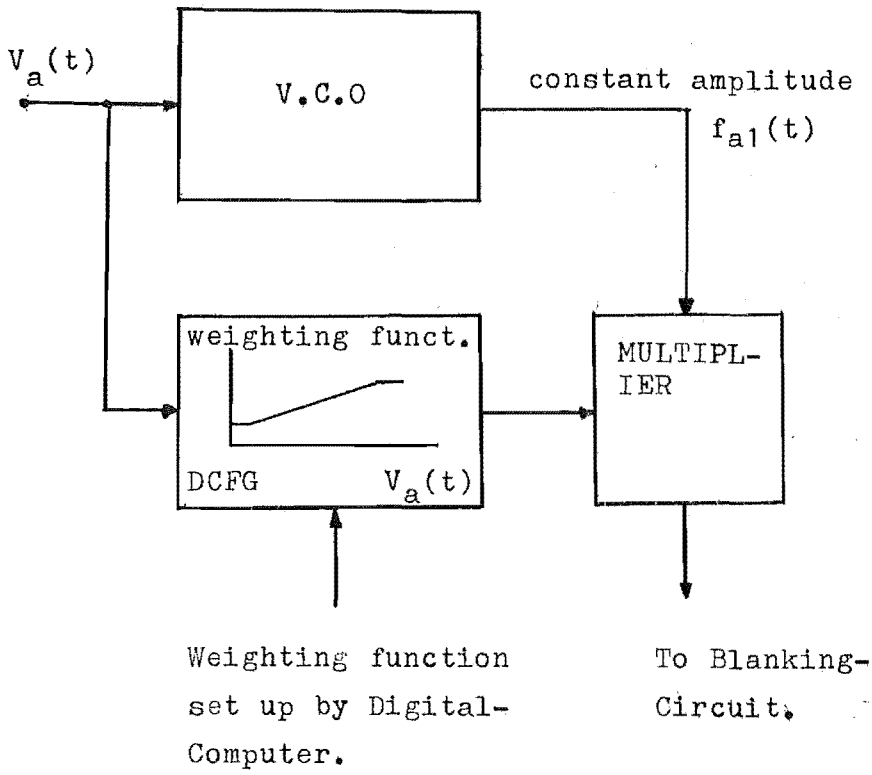


Fig.7. A Set Up to Control the Amplitude of the Descending Tone, $F_{a1}(t)$.

tone as its frequency is descending before it is fed to the blanking-circuit.

Obviously, it is not practical to use a weighting function which reduces the sound intensity as much as, say, 30 dB for each 100 Hz decrease in frequency, since the extent in intensity to which a subject can comfortably listen to should be neither less than 50 dB at 100 Hz, nor more than 90 dB at 3000 Hz. The best form of the weighting function could not be predicted so it was decided to carry out the experiment with various constant slope (db/octave) shaping functions to find out the best possible slope assigned for each frequency.

Three weighting functions having the constant slopes of 4, 6, and 8 dB/oct over a range of 5 octaves from 100 Hz to 3200 Hz used in the experiment (Fig. 6) were generated from the Digital Controlled Function Generator, (DCFG). The voltage function $V_a(t)$ which was used to control the frequency $f_{al}(t)$ was also used to control the gain of the output of the V.C.O. The set-up is shown in Fig. 7.

Since at the time this experiment was carried out, none of the subjects of the previous experiment were available, two other subjects, S_4 and S_5, were chosen. However, in order that the comparison between subjects' performances when with and without weighting-functions became possible, S_4 and S_5 were also seated for the tests similar to that of Experiment 1. Defining the following symbols for the frequency resolutions in four different cases:

$\Delta f(1)$: without weighting function (similar to that of experiment 1)

$\Delta f(4)$: with weighting function of 4 dB/oct slope

$\Delta f(6)$: with weighting function of 6 dB/oct slope

$\Delta f(8)$: with weighting function of 8 dB/oct slope

Tables 4 and 5 show the results obtained by subjects S_4 and S_5 respectively.

TABLE 4 - FREQUENCY RESOLUTION PERFORMANCE OF S_4

f_{a2} Hz	$\Delta f(1)$ Hz	$\Delta f(4)$ Hz	$\Delta f(6)$ Hz	$\Delta f(8)$ Hz
1000	145	140	165	150
2000	405	300	380	445
3000	415	320	390	455

TABLE 5 - FREQUENCY RESOLUTION PERFORMANCE OF S_5

f_{a2} Hz	$\Delta f(1)$ Hz	$\Delta f(4)$ Hz	$\Delta f(6)$ Hz	$\Delta f(8)$ Hz
1000	215	205	185	170
2000	470	440	360	335
3000	585	545	560	580

These results were averaged over different IAD's, where $f_{a1}(t)$ was recorded four times for each combination of f_{a2}, IAD_1 and IAD_2. The standard deviation of Δf is approximately equal to 20 Hz , 55 Hz , and 60 Hz when f_{a2} is equal to 1000, 2000, and 3000 Hz respectively.

Table 4 shows that subject S_4 has the best resolution with a weighting function of constant slope of 4 dB/oct. He also reported that the high tone was louder when the other two weighting functions were used. For subject S_5, the 8 dB/oct curve gave the best improvement at the frequencies 1000 and 2000 Hz. His performance at the frequency of 3000 Hz was not improved very much by any of the weighting

functions used. Perhaps a curve having slope of 8 dB/oct at 1000 Hz and 2000 Hz, and slope of 4 dB/oct at 3000 Hz may give the best performance from this subject.

4.6 DISCUSSION

In 1940, Fletcher [14] introduced the concept of the Critical Band (C.B) to determine the amount of noise which effectively masks a pure tone. Since that time, many experimenters have had the tendency to compare this critical bandwidth with the capability of the ear in analysing the frequency components of a multiple tone stimulus [15,16]. It was thought that both phenomena, the masking of tone by noise and the resolution of tones, could be explained by the same hypothesis of critical bands. However the fact that the ear can easily detect a frequency change of 20 times smaller than the value of the critical band width [17,18], can track a fast time varying tone (as experienced by the sensory aid users), and can resolve the frequency components better in dynamic situations (Table 3) obviously shows that the critical band theory can not explain the capability of resolution of the ear.

During the discussion in chapter 2, it was postulated that it is the combination tones produced by the nonlinearity of the inner ear when two tones are simultaneously sounded, which interfere in the frequency analysation, and worsen the resolution capability. The results of the experiments 1, 2, and 3 obtained in dynamic conditions can be explained with the following modified postulation: The inner ear can perform a real time frequency analysis of the stimuli but sustain a degree of distortion due to its nonlinearity. The distortion products are the combination tones discussed in Chapter 2. The assumption of a real time frequency analyser mechanism for the inner

ear explains why a time varying tone can be tracked. This agrees with Békésy's observation that a travelling wave on the basilar membrane takes only 5 milliseconds to pass from the window to the apex of the cochlea [19]. These are the resonant places of the highest and lowest audible frequencies respectively (noting that a blind aid user approaching an object at about 5 km/hr hears a cyclic variation tone sweeping at a rate of 2.6 kHz/sec only).

The first step in explaining the results obtained in the experiments 1, 2, and 3 is to compare the most prominent subjective components heard when a subject is presented with two tones $f_{a1}(t)$ and f_{a2} simultaneously in the first two experiments (see Table 4).

TABLE 6 - THE MOST PROMINENT SUBJECTIVE COMPONENTS HEARD BY THE SUBJECT IN EXPERIMENTS 1 AND 2

	EXPERIMENT 1	EXPERIMENT 2
High tone, f_{a2}	$f_{a1}(0)$	$f_{a1}(0)$
Low tone, $f_{a1}(t)$	$f_{a1}(0) - Kt - Kt_s$	$f_{a1}(0) - Kt$
Cubic Difference Tone, $2f_{a1}(t) - f_{a2}$	$f_{a1}(0) - 2Kt - 2Kt_s$	$f_{a1}(0) - 2Kt$

where $K = \frac{2v}{c} m$, $0 \leq t_s \leq T_s$ (the repetition period)

It is obvious that as t increases, and t_s varies from 0 to T_s then resets periodically, the saw tooth cyclic variation of $f_{a1}(t)$ in experiment 1 always gives extra instantaneous separations between the subjective components which are not obtainable with the linear variation mode of $f_{a1}(t)$ in experiment 2. Hence the ear can resolve the two components, $f_{a1}(t)$ and f_{a2} , in experiment 1 before it can do so in experiment 2. The same argument applies to the other

combination tones, $(n+1)f_{a1} - nf_{a2}$, and to experiment 3 where the rate of change of $f_{a1}(t)$ is doubled.

The remaining question is why the resolution performance in experiment 2 (linear variation) is better than that in static conditions. Apparently, at the t when the descending tone $f_{a1}(t)$ is separated from the high tone f_{a2} an interval equal to the frequency difference of a pair of stationary tones, the combination tones generated in the case of linear variation and in stationary conditions must be correspondingly placed at the same locations. Hence subject must use other additional cues to enhance his resolution capability in the dynamic case.

It has been seen in Table 6 that while the high tone is stationary, the cubic difference tone descends at a rate of change in frequency equal to twice that of the low fundamental tone $f_{a1}(t)$. This can be generalized that as the stimuli have different rates of change in frequency, all the subjective components will vary at the rates of change in frequency which are entirely different from each other. Hence it is possible that at high level perception, the collicular neurones respond to these subjective components differently, so the interference of the combination tones in the perception of the louder fundamentals is less significant than that in the static experiment when both the fundamentals and the combination tones are of the same characteristics: "STATIONARY" .

Little knowledge about this performance of human beings was found in literature. Su-Ga [19,20,21] studied the responses of collicular neurones of bats to frequency modulated (FM) and complex sounds, classified them in five main types: symmetrical, asymmetrical, FM-insentitive, FM-sensitive, and upper threshold units; and suggested the following principles believed to apply to all kinds of mammals and

human beings but of course with different time scales:

(1) When responding to complex sound, single neurones are most concerned with the components having frequencies identical to their BEST FREQUENCIES, except for FM-sensitive units.

(2) A symmetrical unit responds to any frequency modulation of the best frequency component, and the response is scarcely inhibited by other components of the complex sound.

(3) In an asymmetrical unit, the extent of frequency modulation of the best frequency which can excite the neurone is limited by the inhibitory areas on one or both sides of an excitatory area.

(4) The limitation is stronger for frequency modulation occurring towards the best frequency than for frequency modulation starting from it.

(5) The response to the best frequency component is inhibited by lower and/or higher components unless these are outside the inhibitory area. Neurones sensitive only to downward sweeping tones were more often sampled than those sensitive only to upward sweeping tones.

(6) In an FM-insensitive unit, inhibitory areas on both sides of an excitatory area limit the extent of frequency modulation of the best frequency component which can activate the neurone. This type of neurone responds to more restricted combinations of components than does the symmetrical unit.

(7) An FM-sensitive unit responds selectively to frequency modulation of certain components. The direction, range, rate and functional form of the frequency sweep are especially important in activating the type of neurone. The condition necessary for the excitation depends also on the other components.

(8) An upper threshold unit does not respond to the best frequency component and/or its frequency modulation when its intensity is stronger than the upper threshold.

It is obvious that the characters of selective responding of the three types of neurones asymmetrical, FM-insensitive, and FM-sensitive units - explains the capability of the auditory system in analysing the sound structure of the complex tones. Some upper threshold units having also asymmetrical or FM characteristics can analyse the sound both in frequency and in intensity. Concerning my experiments on auditory resolution so far, I believe that the frequency resolution is so poor in the static conditions because all neurones, having the best frequencies correspondingly equal to the frequencies of the fundamentals and the combination tones, respond to them more or less similarly. Conversely, in the dynamic case, each subjective component modulated at a certain rate will be more responded to by certain groups of neurones and inhibited in the others. This enables subject to discriminate the fundamentals from the combination tones more easily. In other words, the high level perception of the fundamentals is less interfered with by the combination tones. An increase in rate of change of frequency of $f_{al}(t)$ increases the difference in frequency modulation rates of the subjective components, consequently it enhances the resolution performance of a subject.

4.7 SUMMARY OF RESULTS

The important results of frequency-resolution can be summarised as follows:

(1) The capability of frequency resolution, under dynamic conditions is significantly improved when compared with that under stationary conditions.

(2) The frequency resolution, measured in the dynamic case, appears to be independent of the value of IAD_1 and IAD_2 . Rowell [5] when determining the frequency difference between two stationary tones so that each could be localised, also reported this non-relationship; however, he also showed a frequency difference of almost an octave.

(3) In CTFM Sonar, the Doppler effect results in an increase of the rate of change of the range coding frequency by a factor of two, within each sweep cycle. This gives an improvement in the resolution capability (compare experiments 1 and 2).

(4) An increase in the relative velocity of target and sonar also improves the resolution capability of the auditory display system, quantitatively as well as qualitatively (comparing experiments 2 and 3). However, if the target velocity results in too large a rate of change of the range coding frequency, the human reaction time becomes an important factor in the measurement of resolving power and the tracking of objects. Thus, in the sonar for the blind when $T_s = 0.250s$, $V = 1.5$ m/s, and $r_1(0) = 3m$, signals will be resolved cognitively long before reaction can be recorded.

REFERENCES

[1] KAY, L. and DO, M.A. An artificially generated multiple object auditory space for use where vision is impaired. Acustica, Vol.36, 1976, 1-8.

[2] KAY, L. Sonic glasses for the blind - a progress report. A.F.B. Research Bulletin, Vol.25, 1972, p.25.

[3] KAY, L. Sonic glasses for the blind - presentation of evaluation data. A.F.B. Research Bulletin, Vol.26, 1973, p.35.

[4] KAY, L. Toward objective mobility evaluation - some thoughts on a theory. A.F.B. New York, Monograph, 1974.

[5] ROWELL, D. Auditory display of spatial information. Ph.D. Thesis, University of Canterbury, 1970.

[6] SMITH, R.P. Transduction and audible displays for broad band sonar systems. Ph.D. Thesis, University of Canterbury, New Zealand, 1973.

[7] DO, M.A. and KAY, L. Resolution in an artificially generated multiple object auditory space using new auditory sensations. Acustica, Vol.36, 1976, 9-15.

[8] FLETCHER, H. and MUNSON, W.A. Loudness, its definition, measurement and calculation. J. Acoust. Soc.Am., Vol.5, 1933, 82-108.

[9] CHURCHER, B.G., KING, A.T. and DAVIES, H. The minimum percetible of intensity of a pure tone. Phil. Mag., Vol.18, 1934, 927-939.

[10] TREMAINE, H.M. Audio Cyclopedia. Howard W. Sams & Co. Inc., N.Y., 1974, 20-23.

[11] WEGEL, R.L. and LANE, C.F. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev., Vol.23, 1924, 266-285.

[12] FLETCHER, H. Speech and hearing. D. Van Nostrand Company Inc., N.Y., 1929.

[13] FLETCHER, H. Speech and hearing in communication. D. Van Nostrand Company, Inc., N.Y., 1953.

[14] FLETCHER, H. Auditory patterns. Review of Modern Physics. Vol.12, 1940, 47-65.

[15] PLOMP, R. The ear as a frequency analyser. J. Acoust. Soc.Am. Vol.36, 1964, 1628-1636.

[16] PLOMP, R. and LEVELT, W.J.M. Tonal Consanance and critical bandwidth. J.Acoust. Soc.Am., Vol.38, 1965, 548-559.

[17] SHOWER, E.G. and BIDDULPH, R. Differential pitch sensitivity of the ear. J.Acoust. Soc.Am., Vol.3. 1931, 275-287.

[18] HARRIS, J.D. Pitch discrimination. J.Acoust. Soc.Am., Vol.24, 1952, 750-755.

[19] VON BÉKÉSY, G. Experiments in hearing. McGraw-Hill, N.Y. 1960.

[20] SU-GA, N. Analysis of frequency-modulated sounds by auditory neurones of echo-locating bats. J.Physiology. Vol.179, 1965, pp.26-53.

[21] SU-GA, N. Responses of critical auditory neurones to frequency modulated sounds in echo-locating bats. Nature, London. Vol.206, 1965, pp.890-891.

[22] SU-GA, N. Analysis of frequency-modulated and complex sounds by single auditory neurones of bats. J. Physiology. Vol.198, 1968, pp.51-80.

CHAPTER 5

PERCEPTION OF INFORMATION GIVEN BY FINITE TARGETS

CHAPTER 5

PERCEPTION OF INFORMATION GIVEN BY FINITE TARGETS

5.1 INTRODUCTION

The study on the multiple target resolution of the CTFM binaural sonar so far has been carried out with single objects only. Under realistic conditions a target always has a certain size, and is characterized by its shape, hence the corresponding audible signals displayed by the binaural sonar at its two channels must be the band limited tones whose spectra are combined in a specific way in accordance with the shape of the target. New terms like "length coding bandwidth" and "angular width code" (the extents of the frequency and the I.A.D. of a tone in the auditory space) introduced to describe the "dimensions" of a complex tone reflected from a finite target will be the main parameters to be studied in the following experiments.

1) Discrimination between two complex tones given by two finite targets.

2) The effect of the "angular width code" on the discrimination.

3) The effect of the "length coding bandwidth" on the discrimination.

4) Comparing sizes of two finite targets.

It is found that the discrimination between the complex tones is partly dependent on their average IAD's. An increase in the frequency - difference for discrimination of two tones is observed when these tones are of small difference in IAD's; this effect becomes

more pronounced when the "angular width codes" (the extents in IAD) of the tones are increased. The comparison among the results of experiments 1, 3 and the frequency resolution of two single tones reported in the preceding chapter leads to the conclusion that a better frequency discrimination can be obtained with the complex tones provided that the bandwidth of the tone is restricted to a specific value, otherwise the frequency discrimination will be worsened, and finally, experiment 4 will show that the "length coding bandwidth" is a good cue for estimating sizes of finite targets.

5.2 SOME THEORETICAL CONSIDERATIONS

It is frequently known that most of the power scattered from a finite target is scattered from a rather small number of locations, often fewer than half a dozen, which may be thought of as "scattering centers" (an expression attributed to R.E. Kell)[1]. Usually they will be at the ends, edges and other discontinuities. Each center scatters power with a certain magnitude, and with a phase that can be related to a certain reference. However in CTFM binaural sonars (the sonic glass and the fishing sonar), the phase differences among the reflected waves from these scattering centers are suppressed during the blanking time of the final audio-output, consequently only ranges and azimuths of the scattering centers are informed.

Let $R_i(t)$ and $\theta_i(t)$ be respectively the instantaneous range and azimuth of the i-th scattering center, where $i = 1, .., n$, and n is the number of scattering centers. The audible signals given by a finite target, at the two channels, Left and Right, will be:

$$S_L(t) = \sum_{i=1} \gamma_i A_L(\theta_i) \cos [2\pi \int f_{ai}(t)\, dt] \qquad (1)$$

$$S_R(t) = \sum_{i=1}^{n} \gamma_i A_R(\theta_i) \cos [2\pi \int f_{ai}(t)\, dt] \qquad (2)$$

where γ_i , θ_i , and $f_{ai}(t)$ represent the strength, the azimuth and the range coding frequency of the i-th center, respectively. $A_L(\theta)$ and $A_R(\theta)$ are the Left and Right receiver patterns which, as shown by Rowell [2], once having the exponential forms as follows:

$$A_L(\theta) = A \exp [- \frac{(\theta + \alpha)^2}{4 \alpha k}] \qquad (3)$$

$$A_R(\theta) = A \exp [- \frac{(\theta - \alpha)^2}{4 \alpha k}] \qquad (4)$$

will result in a linear relationship between the azimuth angle θ_i and the IAD_i of an object. Thus

$$IAD_i(t) = K \theta_i(t) \qquad (5)$$

where k is the localization constant which can be matched with the splay angle, α , of the transducers. A is the maximum gain of the transducers, and $K = \frac{20 \log_{10} e}{k}$.

As shown in the previous chapters, $f_{ai}(t)$ is determined by

$$f_{ai}(t) = \frac{2m}{c} R_i(t) - f_T(t) \frac{2 \dot{R}_i(t)}{c} \qquad (6)$$

Suppose that $R_c(t)$ is the average range of all the scattering centers (or of the target), then

$$R_i(t) = R_c(t) + \delta R_i(t) \qquad (7)$$

and $$f_{ai}(t) = f_{ac}(t) + \delta f_{ai}(t) \qquad (8)$$

where $$f_{ac}(t) = \frac{2m}{c} R_c(t) - f_T(t) \cdot \frac{2 \dot{R}_c(t)}{c} \qquad (9)$$

and $$\delta f_{ai}(t) = \frac{2m}{c} \delta R_i(t) - f_T(t) \cdot \frac{2 \delta \dot{R}_i(t)}{c} \qquad (10)$$

In order to simplify the equation (10) it will be assumed that within a short duration of time, when a certain task such as

discrimination of targets is carried out, the sizes and shapes of the targets are unchanged. Thus $\delta R_i(t)$ is almost constant and $\delta \dot{R}_i(t) \simeq 0$, so

$$\delta f_{ai} \doteq \frac{2m}{c} \delta R_i \doteq \text{constant} \tag{11}$$

For the dimension of azimuth angle, let $\theta_c(t)$ be the average value of all $\theta_i(t)$. Accordingly, $IAD_c(t)$ is the average of all $IAD_i(t)$.

Figure 1 shows the definitions of the "length", $L(t)$, and "angular width", $M(t)$, of a finite target in real space and the corresponding parameters, the "Length coding band width", $B(t)$, and the "Angular width code", $AWC(t)$, in auditory space.

Figure 2 shows the left and right audio spectra given by a finite target appearing in the left-hand side of the field of view. The difference in locations of the scattering centers results in the difference of the forms of the spectra at two channels. For simplicity, these scattering centers are assumed to have equal strengths.

5.3 DISCRIMINATION BETWEEN TWO COMPLEX TONES GIVEN BY TWO FINITE TARGETS

Considering two finite targets A and B simultaneously presented in the field of view of the sonar and described by their locations and sizes as follows:

Target A : $R_c(t)$, $\theta_c(t)$, $L(t)$, $M(tO$

Target B : $r_c(t)$, $\delta_c(t)$, $l(t)$, $m(t)$

The corresponding variables describing the two complex tones A and B are then:

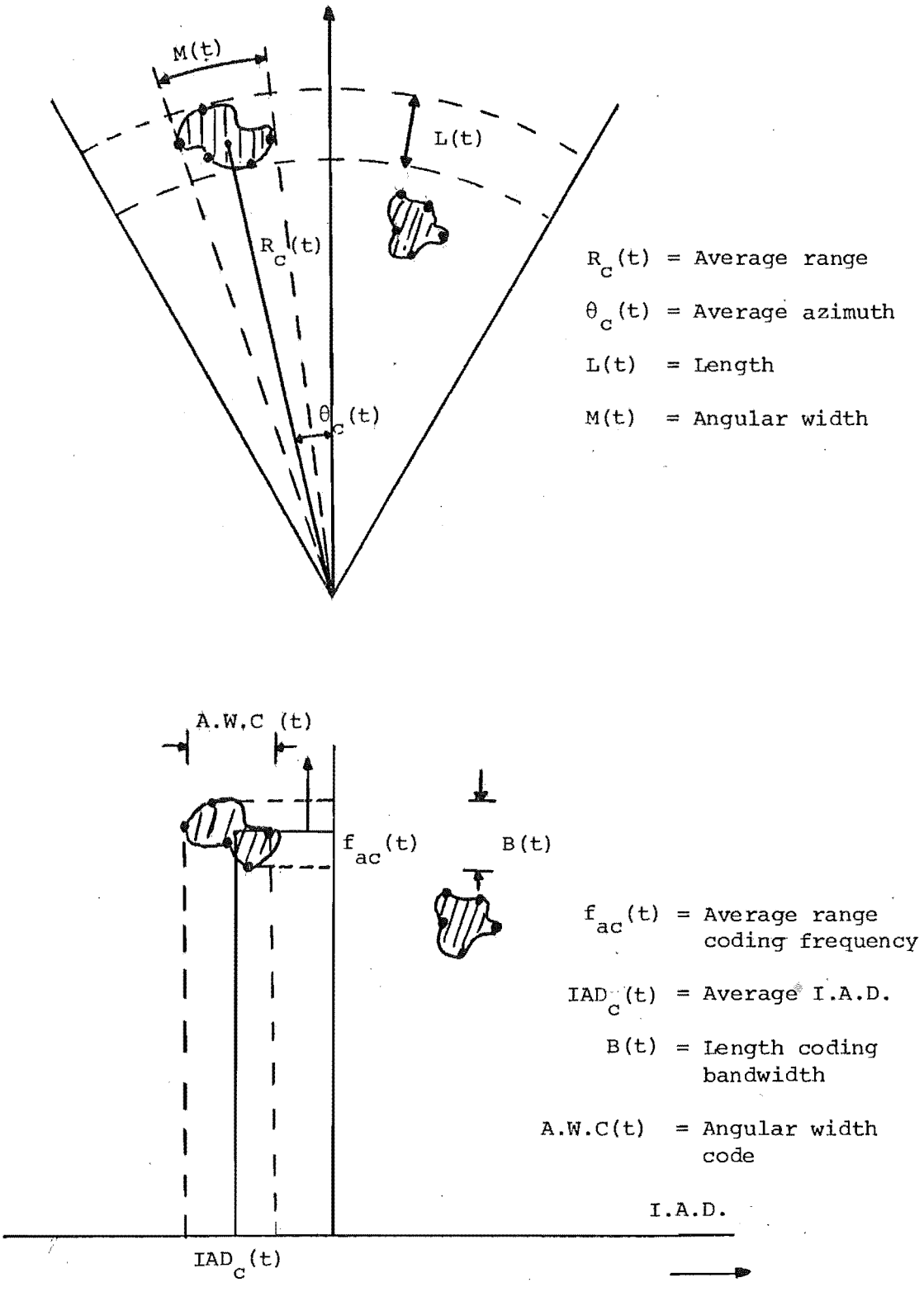


Fig. 1. Relationship between the variables in real space and in auditory space.

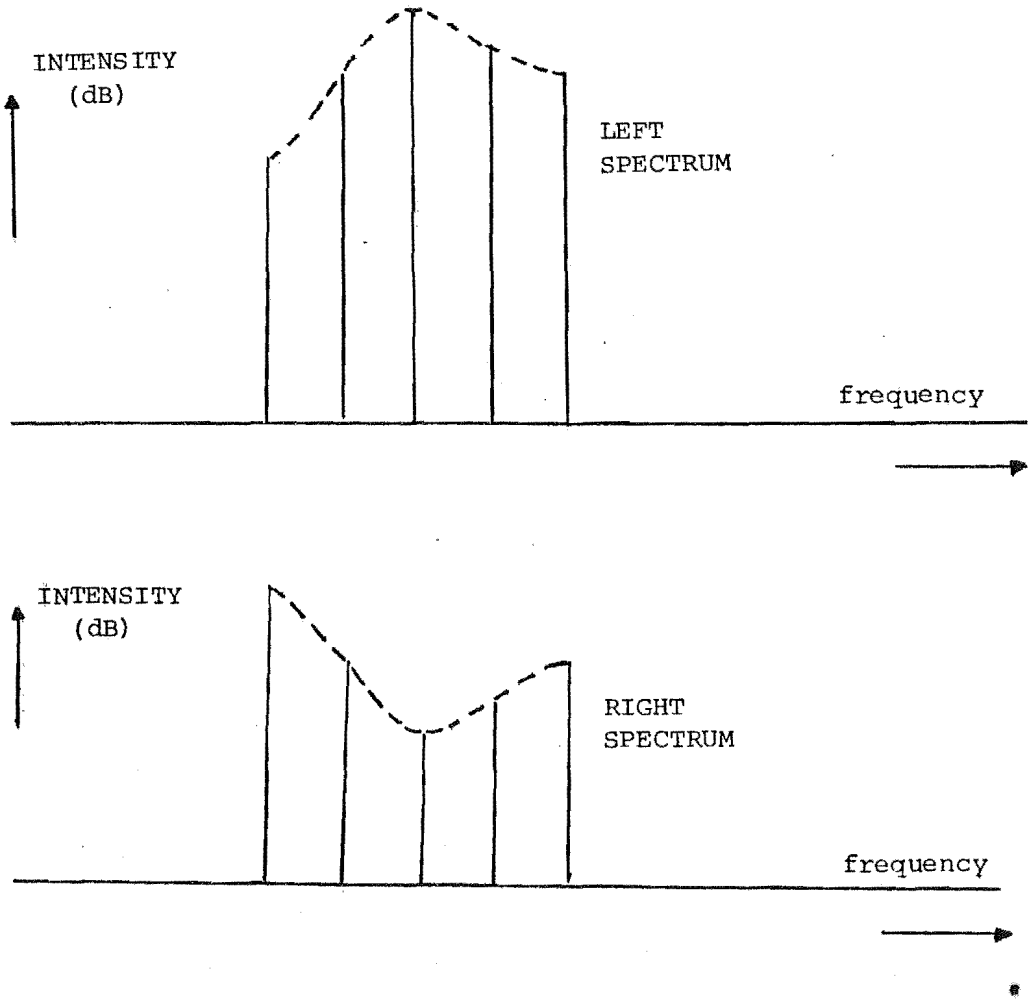


Fig. 2. Left and Right spectra of a complex tone binaurally represented

Tone A : $f_{ac}(t)$, $IAD_c(t)$, $B(t)$, A.W.C (t)

Tone B : $g_{ac}(t)$, $IAD_c(t)$, $\beta(t)$, $\alpha.\omega.\gamma$ (t)

As mentioned in the previous section, it is assumed that during the time the discrimination task is carried out, the changes in size of the targets are not significant. So to simplify the problem, the length coding bandwidths and the angular width codes will be kept as constant parameters in the experiments. The function describing the discrimination of two tones is then reduced to a function of four variables $(f_{ac}(t), IAD_c(t), g_{ac}(t), IAD_c(t))$. This function depends on the similar variables as that of the pure-tone resolution. Hence in the experiment to be conducted, g_{ac}, IAD_c and JAD_c are held constant, varying only $f_{ac}(t)$ as described in equation (9).

5.3.1 The Simulation

Only the simplest case of constant approach velocity, v, is considered, therefore $f_{ac}(t)$ can be put under the forms.

$$f_{ac}(t) = f_{ac}(o) - \frac{2v}{c} m\, t - \frac{2v}{c} m\, t_s \qquad (12)$$

where $o \leq t_s \leq T_s$.

The control voltage required to generate that frequency is

$$V_{ac}(t) = V_{ac}(o) - \alpha\, t - \alpha\, t_s \qquad (13)$$

where $V_{ac}(o)$ and α are linearly related to $f_{ac}(o)$ and $\frac{2v}{c} m$.

A ten-channel voltage controlled oscillator was built to synthesize two complex tones, each of five components. The five components of the moving tone were $f_{ai}(t) = f_{ac}(t) + \delta f_{ai}$ where $i = 1,\ldots, 5$; $\delta f_{al} = 0$ while the other $\delta f_{ai} (i \neq 1)$ could be either positive or negative. The corresponding control-voltages

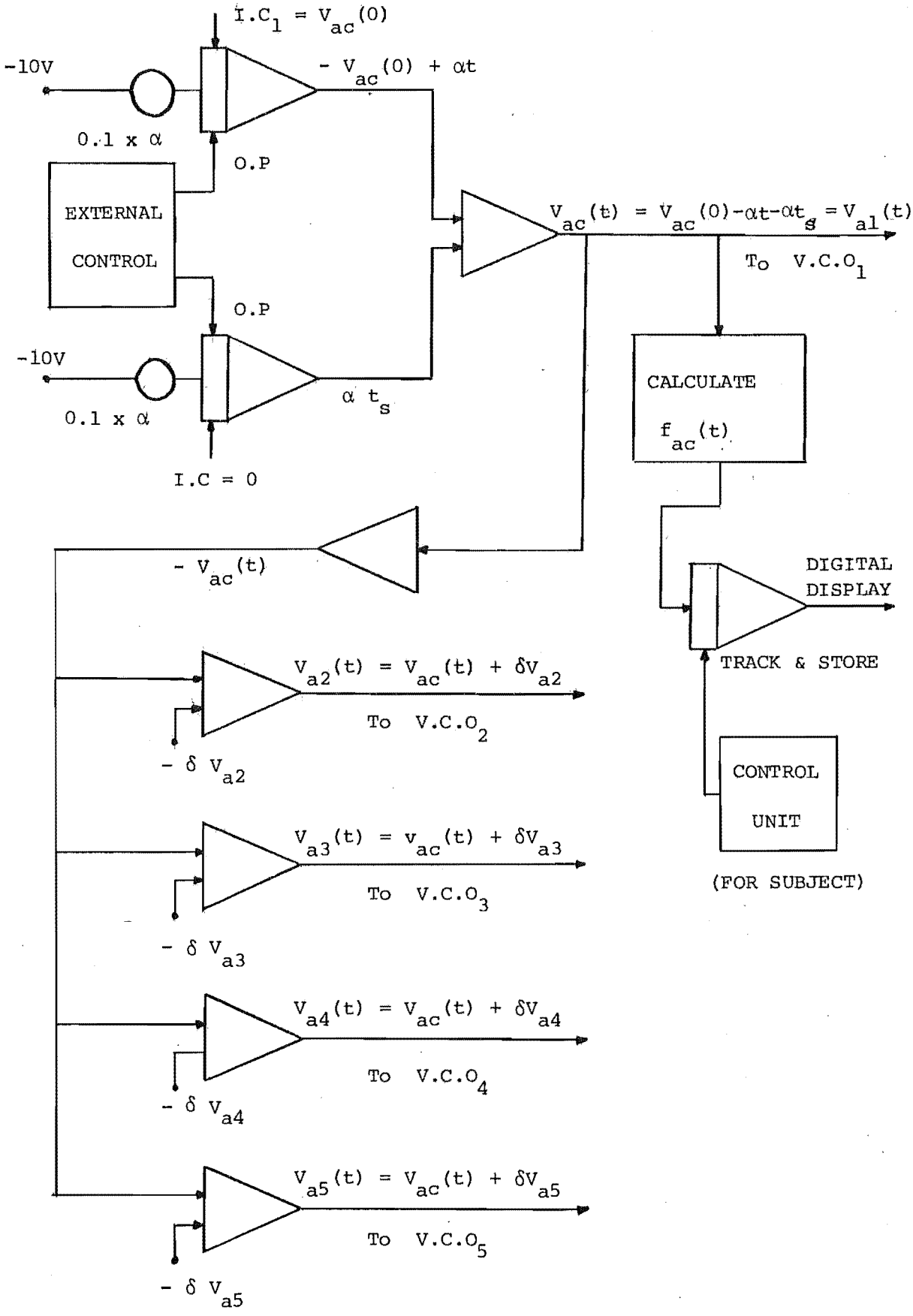


Fig. 3. Analogue computer to generate $V_{ai}(t)$, i = 1,5; and to control experiments.

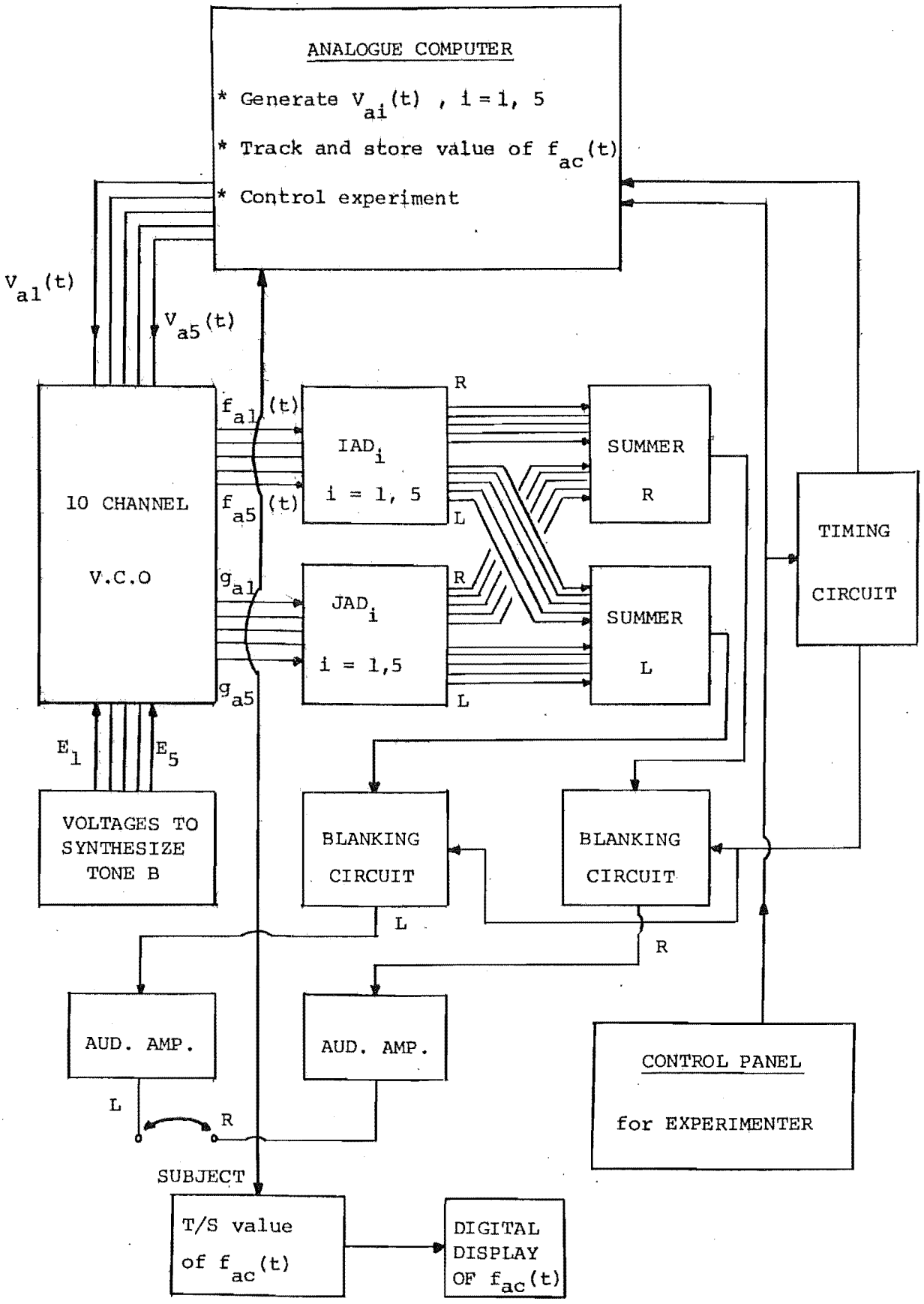


Fig. 4. Arrangement of equipment for the experiments.

were $V_{ai}(t) = V_{ac}(t) + \delta V_{ai}$, where $\delta V_{a1} = 0$, $\delta V_{ai}(i \neq 1)$ could be either positive or negative. An analogue computer was used to produce the voltage functions $V_{ai}(t)$, to compute and to track and store the value of $f_{ac}(t)$, and to control the experiment.

Figure 3 shows the panel patching of the computer, and Figure 4 shows the arrangement of the equipment for the experiment.

5.3.2 Experimental Procedure

The procedure of the experiment was based upon the psychological phenomenon similar to that found in the experiment with pure tones. The discrimination between two complex tones is first obtained at the instant that the two tones separate and slide in auditory space to their respective positions (Figure 5).

The subject's task was to listen to the complex sound including $f_{ac}(t)$ and g_{ac} , where $f_{ac}(o) \simeq g_{ac}$, as described and press the switch to store the value of $f_{ac}(t)$ at the instant the descending tone was perceived to slide to the left or the right of the initial combined image. Each subject was questioned about the direction of shift of either the low tone, $f_{ac}(t)$, or the high tone, g_{ac} . For each value of g_{ac} , the experimenter recorded the values of $f_{ac}(t)$ for different values of IAD_c and JAD_c . These latter are changed in a random manner according to a table.

5.3.3 Experiment 1. (Complex tones of narrow band width and small extent in IAD .)

The simulations relate to the fish sonar where the approach velocity of target A is $v = 1.5$ m/s , the sound velocity is taken as 1500 m/s , $T_s = 3.2$ sec , and $m = 12.5$ kHz/sec . Then

$$f_{ac}(t) = f_{ac}(o) - 25\ (t + t_s) \qquad (14)$$

$$0 \leq t_s \leq 3.2 \text{ s} .$$

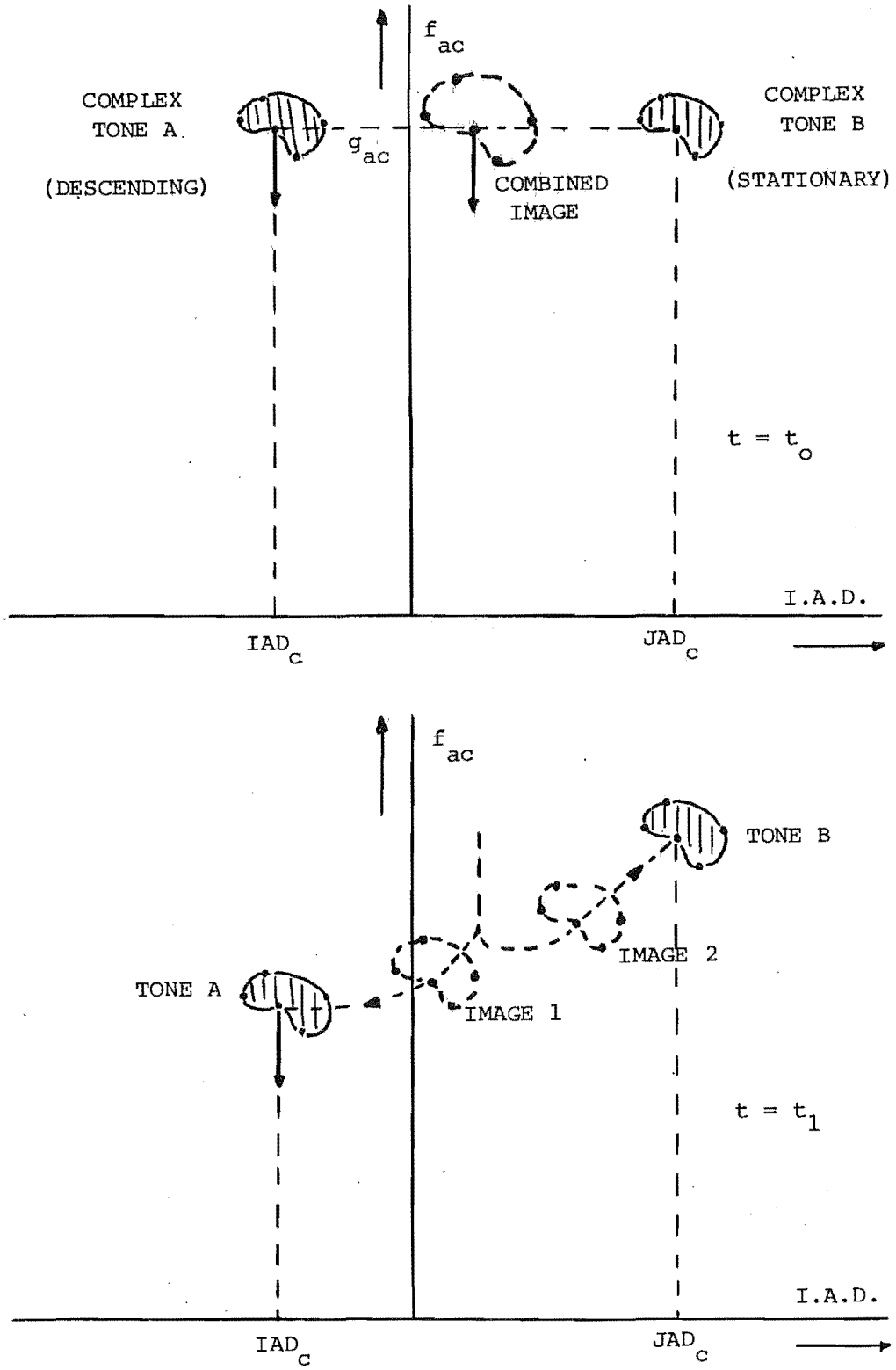


Fig. 5. The phenomenon of discrimination of two complex tones.

The "dimensions" of the complex tones and the corresponding sizes of the targets are as follows:

Length coding bandwidth = 100 Hz	target length = 6 m
Angular width code = 2dB	angular width = 4 deg.

The components of each tone are equally spaced.

Three subjects S_1 S_2 S_3 were used in this experiment. Values of $f_{ac}(o) = g_{ac}$ were chosen to be 1000, 2000 and 3000 Hz, while the values of IAD_c were 0, + 4, + 8 dB for the descending tone and IAD_c were equal to -10, -6, -2, +2, +6 and +10 dB for the stationary tone. Similar to the experiment with pure tone, an error of up to - 25 Hz or + 55 Hz may be introduced into the recorded frequency discrimination due to the delay of the subject's reaction. For each combination of g_{ac}, IAD_c and JAD_c, the recording of the frequency discrimination in four times shows a relatively small deviation, frequently less than 20 Hz.

Figure 6a, b, c shows the variation of frequency-discrimination performed by S_1 with respect to frequency, at different values of IAD_c and JAD_c. The results given by subjects S_2 and S_3 are very close to this (see Appendix 2), an example is shown in Figure 7. Besides the better discrimination performance obtained in this case of complex tone in comparison to the performance with pure tones, it was noted that the frequency discrimination was slightly altered when IAD_c and JAD_c are closed together. Although this increase in Δf_{ac} was not very large, it was consistently found in the performances of all three subjects at different frequencies, and IAD's. To recognise this effect easily, the frequency-discrimination was rearranged as a function of JAD_c while g_{ac} and IAD_c are held as constant parameters (Figure 8 a to i).

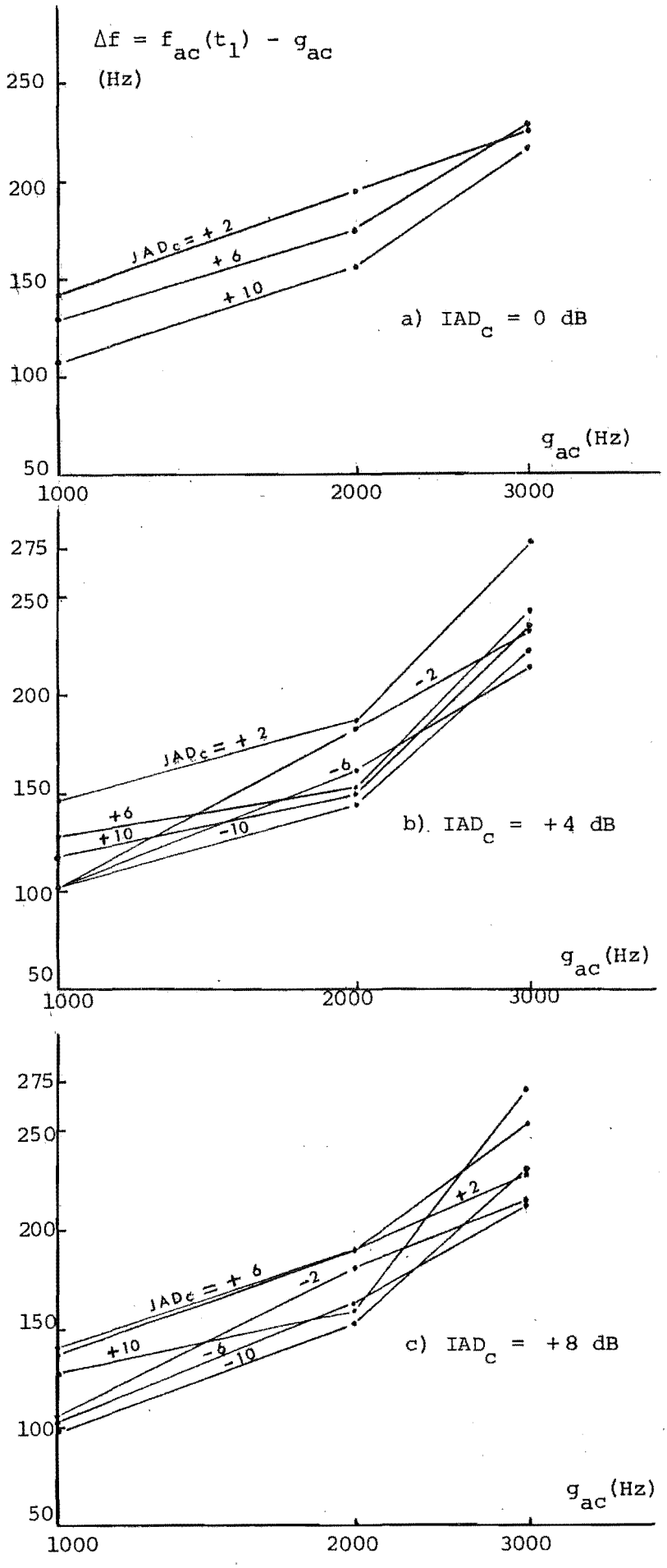


Fig. 6. Frequency discrimination performed by S_1, presented as a function of frequency.

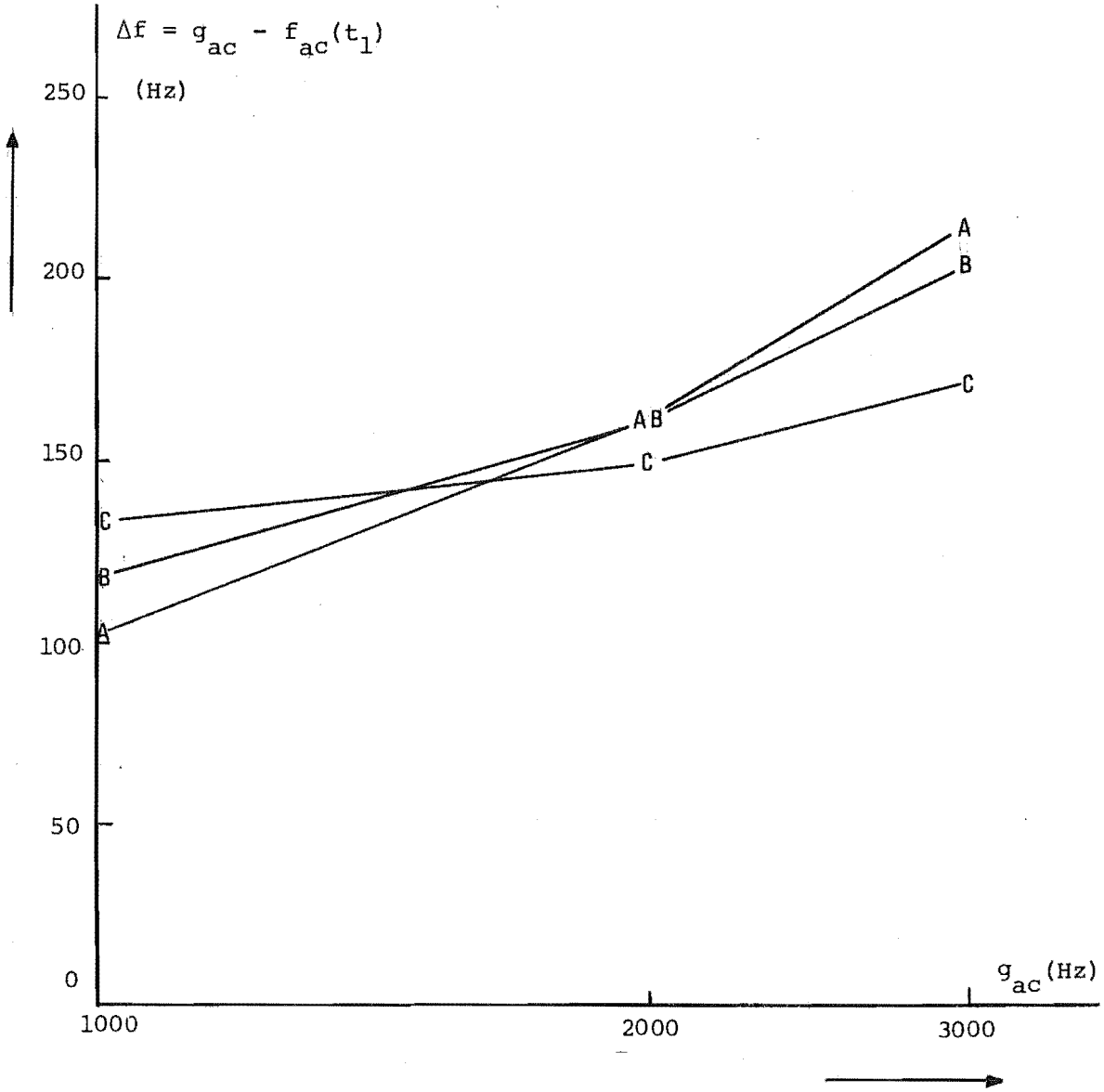


Fig. 7. Comparison of frequency discrimination performances of three subjects, $S_1 \equiv A$, $S_2 \equiv B$, $S_3 \equiv C$, in the case of $IAD_c = +4dB$, $IAD_c = -6$ dB.

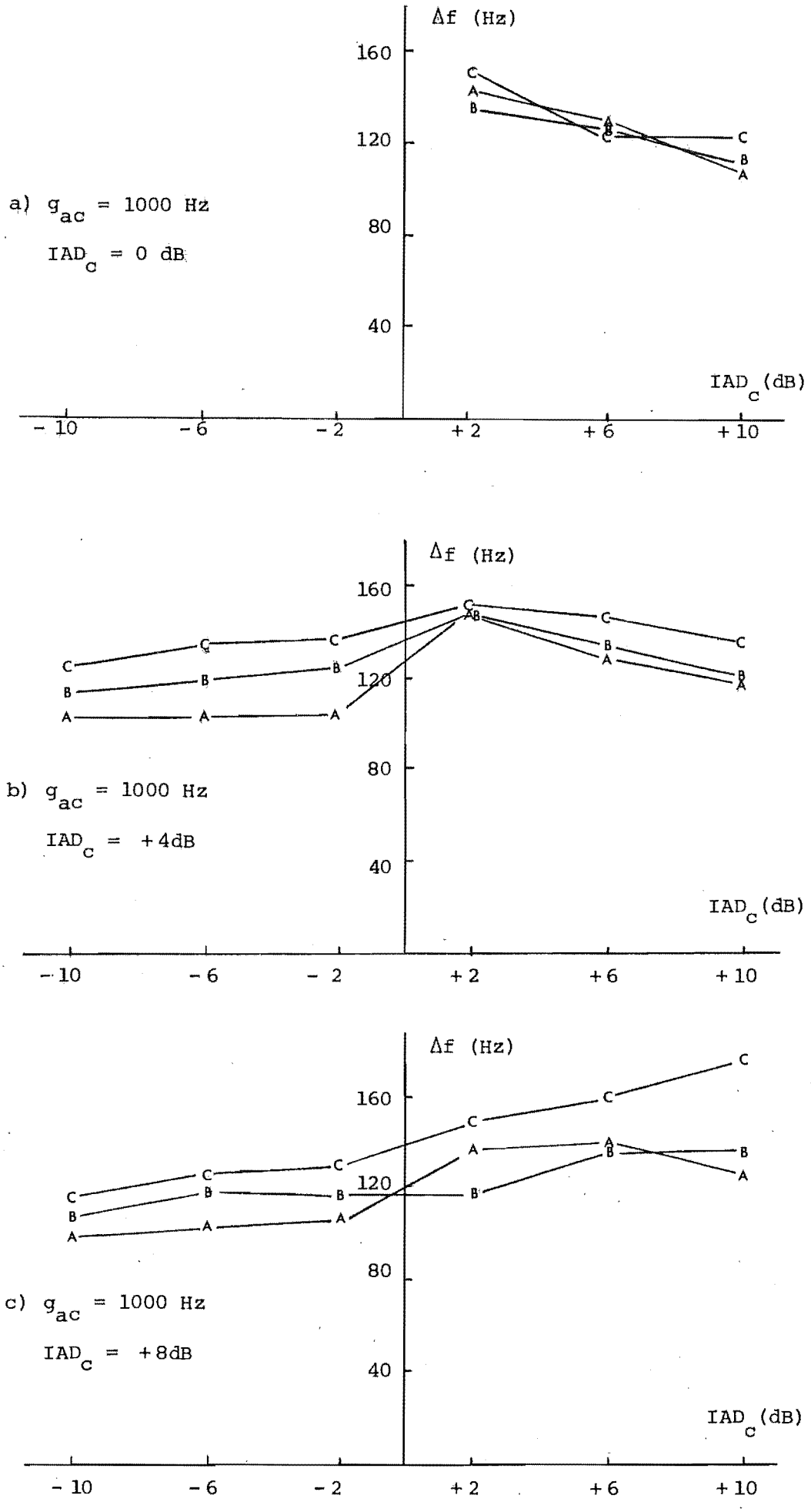
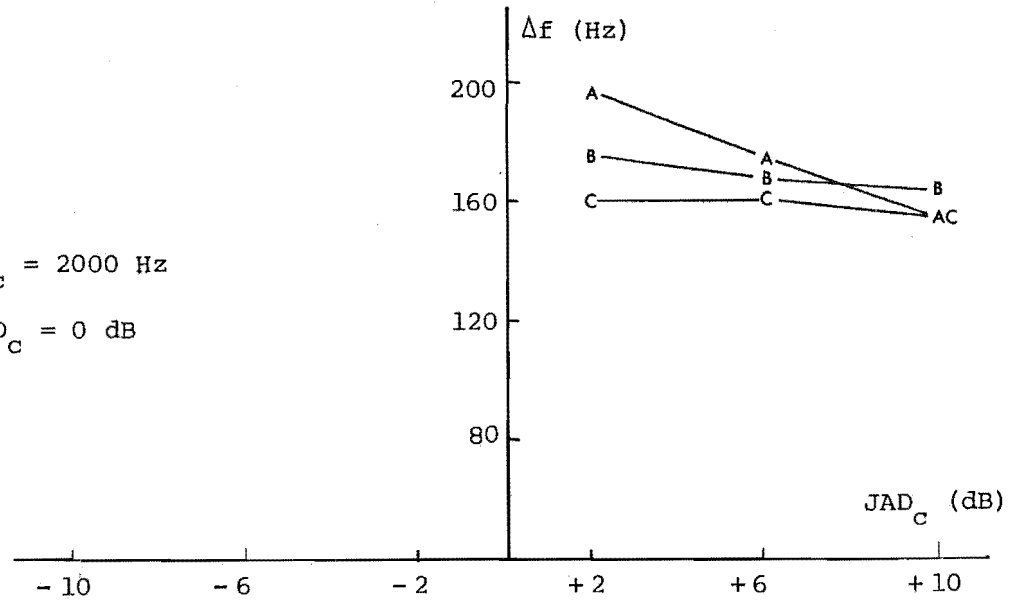


Fig. 8. Frequency discrimination as a function of I.A.D. performed by three subjects.

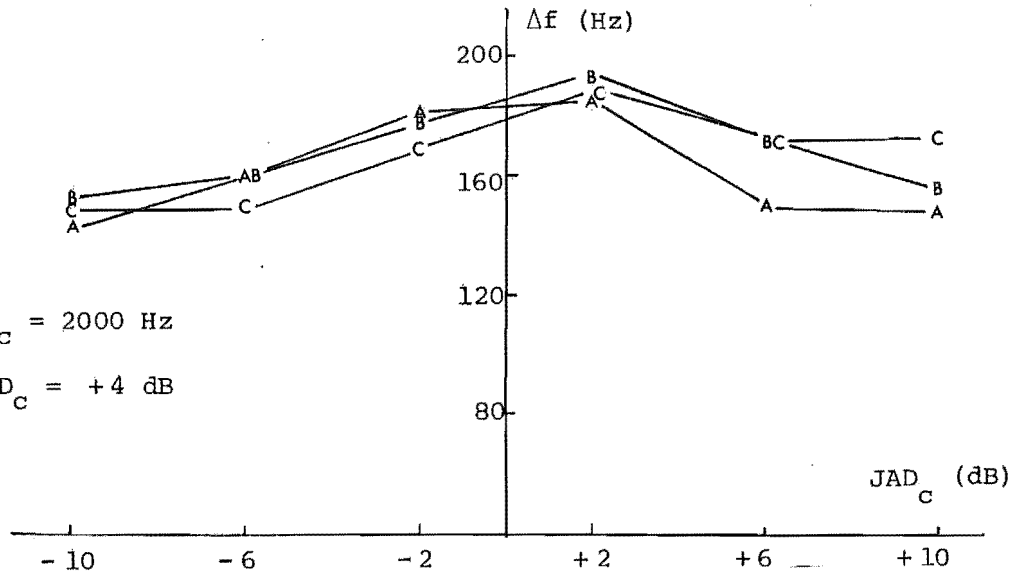
d) g_{ac} = 2000 Hz

IAD_c = 0 dB



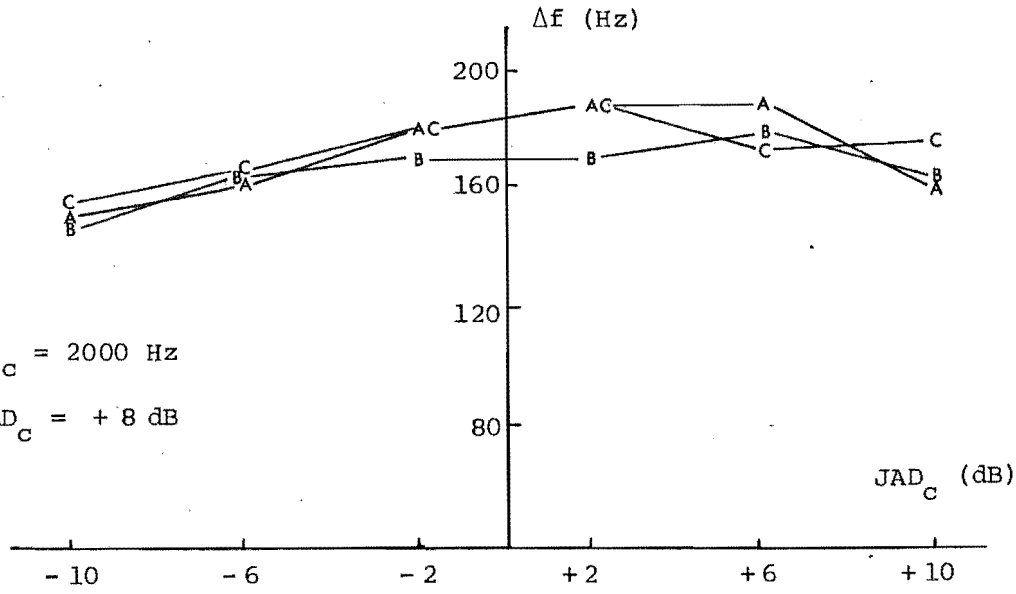
e) g_{ac} = 2000 Hz

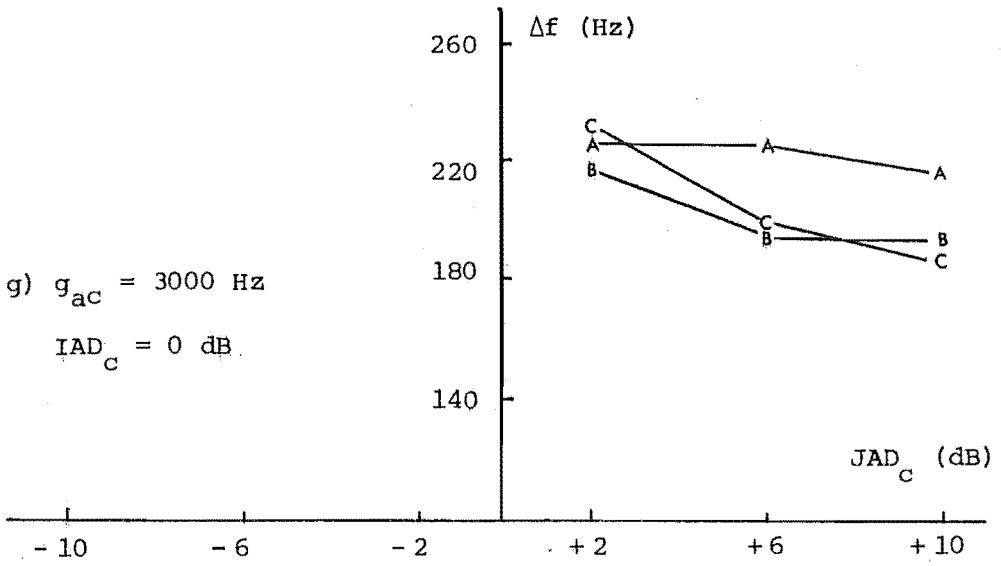
IAD_c = +4 dB



f) g_{ac} = 2000 Hz

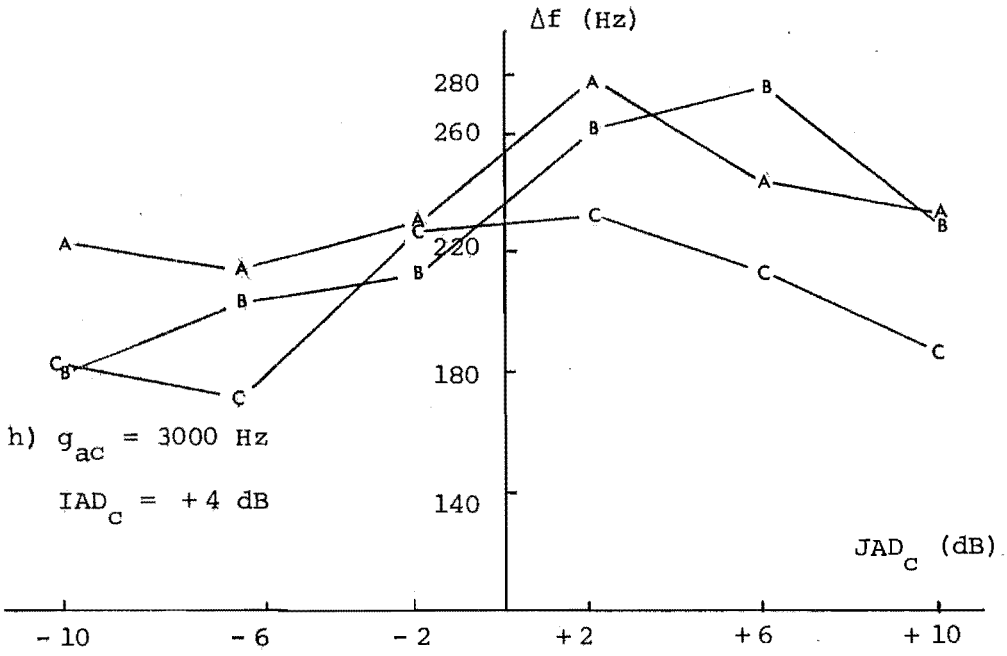
IAD_c = + 8 dB





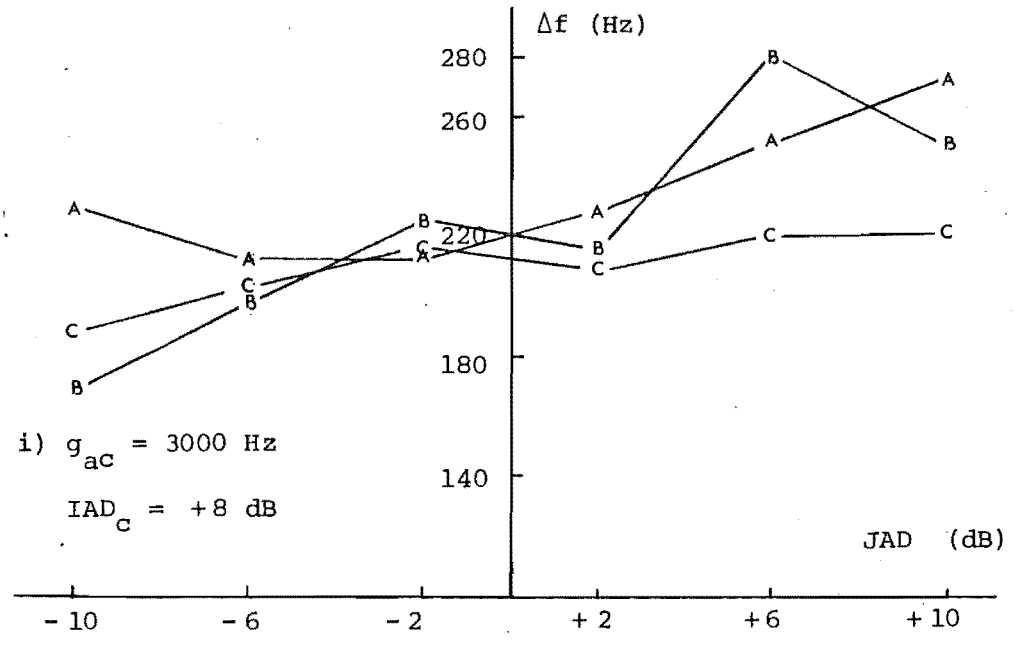
g) g_{ac} = 3000 Hz

IAD_C = 0 dB



h) g_{ac} = 3000 Hz

IAD_C = + 4 dB



i) g_{ac} = 3000 Hz

IAD_C = + 8 dB

5.4 THE EFFECT OF ANGULAR WIDTH CODE ON FREQUENCY DISCRIMINATION. (Experiment 2: Complex tones of narrow bandwidth but large extent in IAD.)

In order to study the effect of the angular width code on frequency discrimination, the extents in IAD of the tones were increased to 4dB, and a similar experiment was repeated with subject S_2 (this subject is the same subject S_2 used in the experiment with pure tones). (A 4dB extent in IAD of the complex tone is equivalent to an 8 deg. angular width of the finite target.) Results are presented in Appendix 2.

Figure 9a, b, c represents the discrimination as a function of JAD_c, while g_{ac} and IAD_c are constant parameters. Again, it is noted that there is a certain increase in Δf_{ac} when JAD_c and IAD_c are less different. A comparison between the performances of subject S_2 in two experiments, 1 and 2, is shown in Figure 10a, b, c. Besides the expected increase of Δf_{ac} when the angular width code is doubled, the increase in Δf_{ac} when JAD_c approaches the value of IAD_c appears to begin sooner, and finish later. In other words, the peak representing this effect is broader when the "angular width code" is larger.

5.5 THE EFFECT OF LENGTH CODING BANDWIDTH ON FREQUENCY DISCRIMINATION. (Experiment 3: Complex tones of large bandwidth but small extent in IAD.)

The length coding bandwidth of the complex tone is now doubled and equal to 200 Hz, while the angular width code is restored to the value of 2dB. Since the effect of IAD on frequency discrimination has been understood through experiments 1 and 2, only five combinations of IAD_c and JAD_c are involved in this experiment.

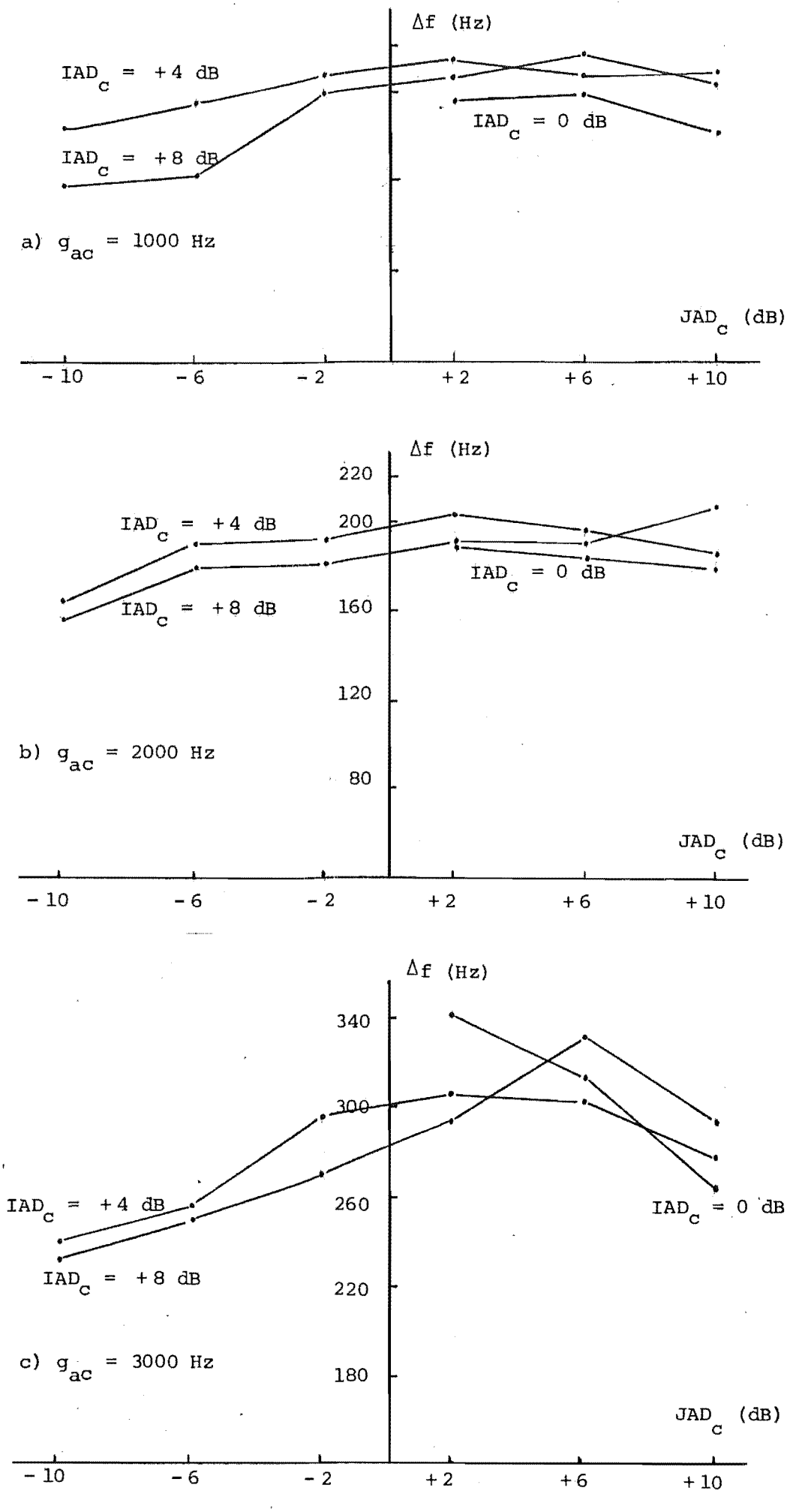


Fig. 9. Frequency discrimination performed by subject S_2 with complex tones of large extent in I.A.D. (A.W.C. = 4 dB).

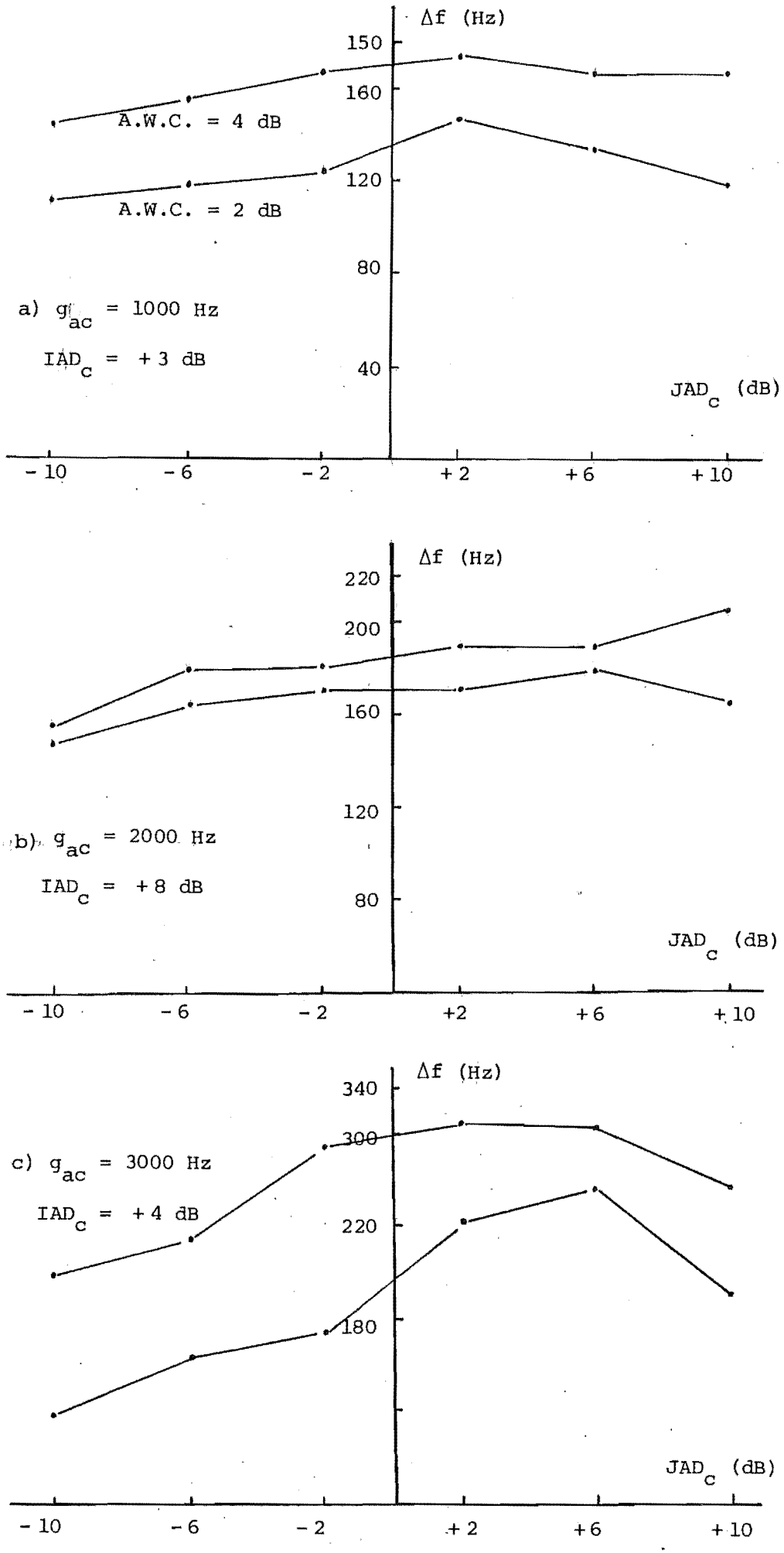


Fig. 10: Comparison of performances of subject S_2 in two cases of narrow and large extents in I.A.D.

This number of combinations is still large enough for the experimenter to select them randomly so that the subject cannot give a "Left" or "Right" judgement of a tone by guessing or remembering the initial pattern (if only two combinations of IAD_c and JAD_c are used, subject may remember the initial patterns and guess the direction of the tones before he actually resolves them). The average "frequency discrimination" is presented in Table 1.

TABLE 1. FREQUENCY DISCRIMINATION PERFORMED BY SUBJECT S_2 IN THE EXPERIMENT WITH COMPLEX TONES OF LARGE BANDWIDTH BUT SMALL EXTENT IN IAD (B = 200 Hz, A.W.C. = 2 dB)

g_{ac} Hz	$IAD_c = 0$ $JAD_c = +6$ dB	$IAD_c = +4$ $JAD_c = -6$ dB	$IAD_c = +4$ $JAD_c = +6$ dB	$IAD_c = +8$ $JAD_c = -6$ dB	$IAD_c = +8$ $JAD_c = +6$ dB
1000	210	235	260	200	230
2000	265	240	300	250	355
3000	350	370	420	360	425

The comparison of this result with the corresponding result obtained in experiment 1 shows that there is, on average, an increase of Δf_{ac} equal to 79% at g_{ac} = 1000 Hz, 65% at g_{ac} = 2000 Hz, and 69% at g_{ac} = 3000 Hz. For the case of g_{ac} = 1000 Hz and length-coding-bandwidth = 200 Hz , subject reported the difficulty in separating the two tones, and larger deviation from the mean frequency-discrimination was recorded, about twice of that in experiment 1 (40 Hz) . The performances of subject S_2 in three cases: (i) resolving two single tones [3], (ii) discriminating two complex tones of narrow bandwidth (100 Hz), (iii) discriminating two

complex tones of large bandwidth (200 Hz), are compared in Figure 11. This comparison seems to lead to the conclusion that "the frequency discrimination is improved for narrow-band tones but worsened for wide-band tones". However, to answer precisely how narrow or how wide the bandwidths of the tones should be so that improving or worsening the frequency discrimination occurs, more elaborate experiments with several subjects are necessary since that work is similar to determining two kinds of threshold at the same time: the frequency discrimination and "the critical bandwidth" (the term "critical bandwidth" is used in a specific meaning related to the frequency discrimination of complex tones in this case, but whether it is or it is not similar to the Fletcher's original concept will be discussed later), while it has been known that determining merely the critical band of noise which effectively masks a tone was reported diversely by several authors [4 - 9]. Alternatively, in another approach, it will be shown that the investigation on the mechanism why such improvement or worsening of frequency discrimination occurs can also reveal some knowledge about the conditions at which that effect takes place.

5.6 AN EXPLANATION FOR THE HIGH FREQUENCY DISCRIMINATION PERFORMANCE WITH NARROW-BAND TONES

It was known from place theory that if the stimulus consists of several frequency components, which are not separately resolved, then an averaging mechanism does take place [10 - 11], a pitch of frequency approximately equal to that of the center component will be heard. An explanation for the improvement of frequency discrimination of narrow-band tones can be based upon that averaging mechanism as follows.

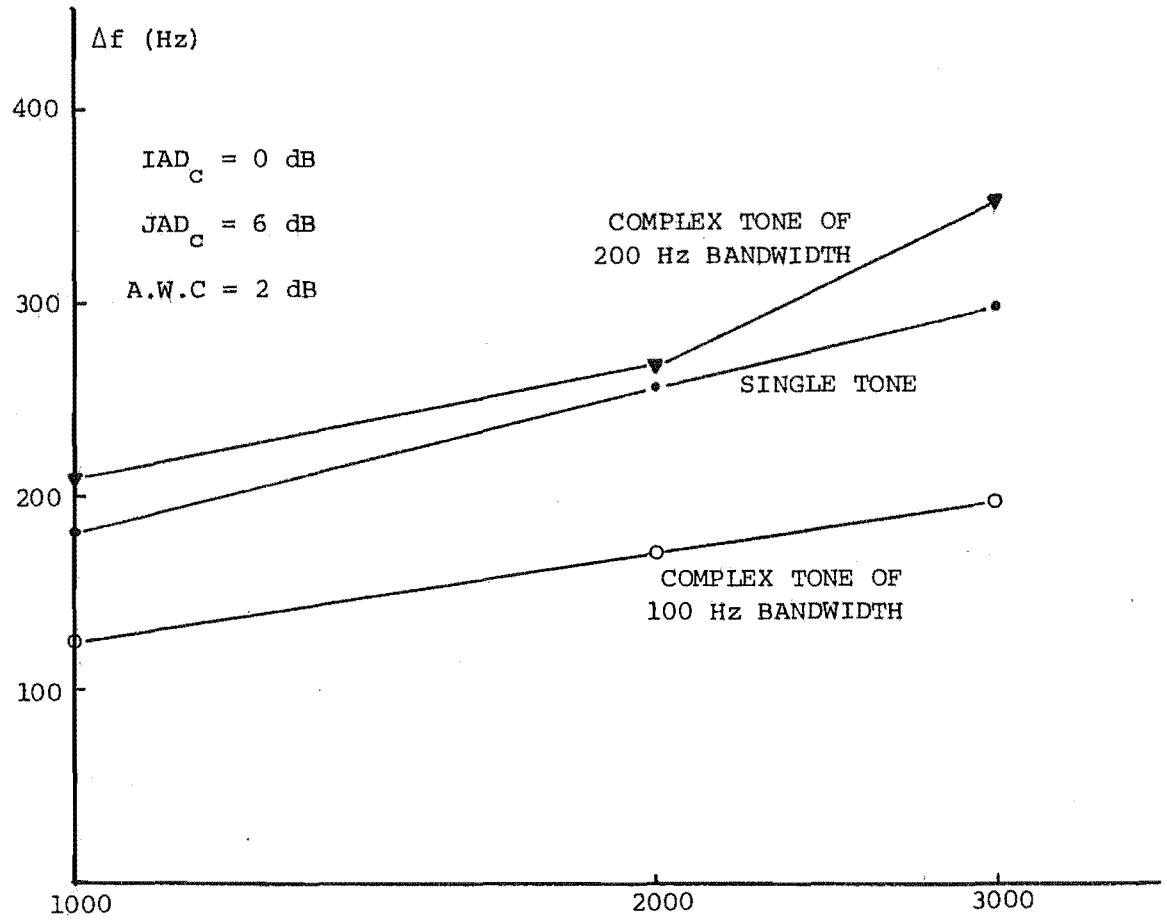


Fig. 11. Comparison of frequency discrimination performances of subject S_2 in three cases: (i) single tone, (ii) complex tone of 100 Hz bandwidth, and (iii) complex tone of 200 Hz bandwidth .

Throughout the discussions in Chapters 2 and 5, it was seen that the CT's, especially the cubic difference tone $2f_1 - f_2$, generated due to non-linearity of the inner ear, interfere significantly in the discrimination of the fundamentals. Figure 12a shows the typical level of the $2f_1 - f_2$ tone when f_1 and f_2 are equal to 70 dB [12 - 14]. In comparison to that, if two complex tones of the same levels (70 dB) as the single tones, and having the bandwidths B's, are considered the cubic difference tones generated by the combinations of various pairs of components will be spread out over a large frequency band as demonstrated in Figure 12b. This also applies to the higher order CT's. Therefore, if the bandwidths of the complex tones are narrow enough so that the averaging process can take place, two stimuli must raise two pitches of frequencies equal to the center frequencies of the complex tones and of intensities equal to that stimulated, while all the CT's are spread out and create a low-level wide band background. The interference of this background may not be as considerable as the interference of the CT's generated by a two pure tone stimulus.

Accepting the above explanation for the improvement of frequency discrimination of narrow band (100 Hz) tones in experiment 1 leads to the only possible explanation for the worsening of frequency discrimination of wide band (200 Hz) tones in experiment 3; that the averaging mechanism fails to summate the components of the complex tones effectively in the later case, the stimulus tones are diffused and the CT's generated by the high complex tone interfere seriously on the low tone (Figure 12c). Previous studies on the loudness summation of complex tones carried out by Zwicker et al. [15,16] and Scharf [17,18] support this hypothesis.

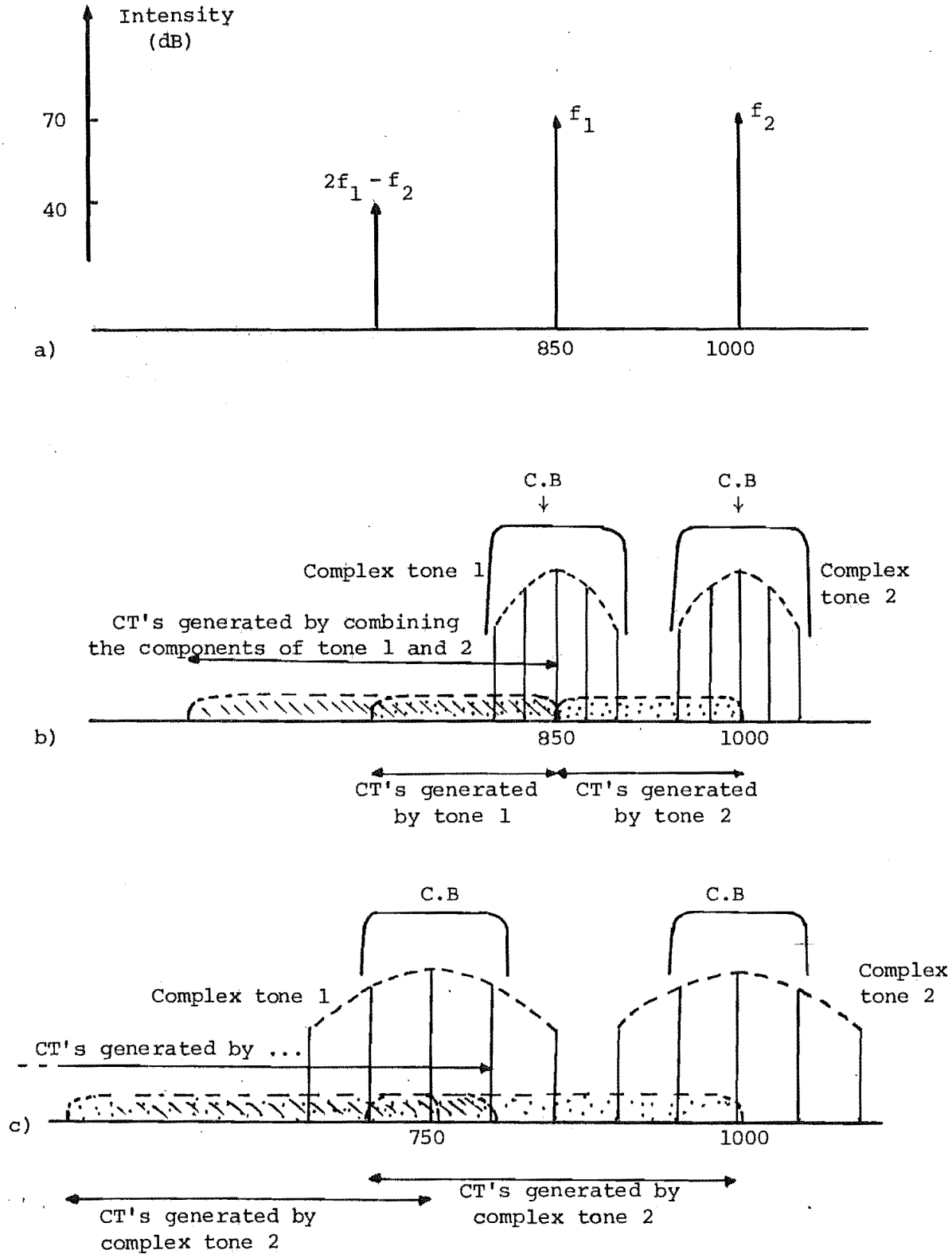


Fig. 12. Explanation for the improvement of frequency discrimination of narrow band tones and the worsening of frequency discrimination of wide band tones

Indeed, according to their results:

(i) Any complex sound having its bandwidth smaller than a critical band (C.B) has a loudness independent from the value of the bandwidth. This critical band is about 2.5 times wider than the critical band described by Fletcher [4] and is the same as those measured in experiments on phase, threshold and two tone masking [15,16] (Figure 13).

(ii) At the sensation level above 15 dB, the spreading of energy in the complex tone over more than a C.B increases the loudness [15 - 17].

(iii) These critical bandwidths are independent from the sound level and the number of components of the complex tone [17 - 18].

What is found in experiment 3 with the complex tones of 200 Hz bandwidths can be explained as follows:

a) At the frequency of 1000 Hz, the bandwidths of the tones (200 Hz) are larger than the C.B's, the tones are diffused and the average mechanism fails to take place, also the CT's generated due to the combining of different components of the high tone interferes on the low tone seriously. Therefore subject S_2 reports difficulty in separating two tones.

b) At the frequencies of 2000 Hz and 3000 Hz, the average mechanism may not fail to take place since the bandwidths of the tones are narrower than the C.B's, but the CT's generated by the high tone still interfere on the low tone much more than they do in experiment 1. The frequency discrimination is worsened but the tones are still easily perceived and separated.

Hence to obtain a high discrimination performance with complex tones, two criteria should be considered:

(i) The bandwidths of the tones must be smaller than the C.B's so that the tones are not diffused and the averaging mechanism takes place.

(ii) If the bandwidths of the tones are still large (even though they are smaller than the C.B's) the CT's generated by the components of the high tone can interfere on the low tone, and the frequency discrimination between the two tones is worsened.

5.7 COMPARING SIZES OF TWO FINITE TARGETS (Experiment 4)

One of the most common methods used in many echo ranging systems to estimate the size of a target is measuring the "reflecting power" of the echo from the target. Terms like "target strength" and "back-scattering cross-section" are well known in this technique. However, estimating the target size in this way is valid only if all the acoustical characteristics of the detected target are well known beforehand or only one kind of target is encountered when the system is in use. In the case of the sonic glasses, the sizes of a bush and a bus-stop post cannot be compared by that method. More clues are needed to discriminate between a large shoal of small fish and a small shoal of large fish in the fishing sonar. The problem is obviously very complicated. The purpose of this experiment is limited to some simple tests to show "qualitatively" how information such as "the length coding bandwidth" and "the angular width code" can be perceived by a subject as the "dimensions" of a complex tone which corresponds to the size of a finite target.

5.7.1 Perceiving Information on the Bandwidth of the Tone

The stimuli are two tones having the same angular width codes of 2 dB. But tone A has a length coding bandwidth of 200 Hz, and

tone B is of 100 Hz bandwidth.

- Initially, both tones had the same center frequencies of 2000 Hz. They were presented successively to subjects at random order and random IAD. Two subjects always reported that the wide-band tone was more "complex", while the experienced third subject recognised immediately that the tone A had a wider bandwidth.
- Two tones were then presented simultaneously, and tone A descended as in previous experiments. When they were separated in auditory space, two subjects could recognise that the wider band tone was moving down in every trial while the other subject needed to listen again a few times before he could do so. To verify a subject's judgement in this experiment, the experimenter had explained to subjects prior to the test that either the wide band tone or the narrow band tone could randomly be the descending tone in either direction (left or right). The truth that only the wide band tone moved down was found by all subjects.

5.7.2 Perception of Information on the Extent in I.A.D. of a Tone

The stimuli are two tones having the same length coding bandwidth of 100 Hz, but one has an angular width code of 2 dB while that of the other is equal to 4 dB. The angular width codes of the tones can be randomly exchanged.

- Initially, both tones had the same center frequencies of 2000 Hz. They were randomly presented to subjects at random order and at different average IAD's. Two subjects could always find the difference in two tones and described the tone of 4 dB angular width code as diffused in auditory space. The third subject required an increase of the angular width code up to 5 dB to recognise this effect.

- Two tones were then presented simultaneously, and tone A swept down as in previous tests. When they were separated, only one subject could recognise which tone was more diffused in direction. The other two subjects could not do this even though the angular width code of the more extended in I.A.D. was increased to 6 dB.

5.8 CONCLUSION

Throughout the above experiment it has been seen that perception of complex tones was accompanied by several characteristics differing considerably from that found in the perception of single tones. The most important result is that discrimination of two complex tones is easier than discrimination of two single tones provided that the bandwidths of the complex tones are restricted to some specific values so that the components of each tone can be summated and averaged out while the interfering CT's are spread out. Differing from the case of single tone, where the frequency discrimination does not depend upon the IAD's of the tones, the diffusions in I.A.D. of the complex tones worsen the frequency discrimination slightly, as the tones are closed in direction. As the extents in IAD of the complex tones are increased, the frequency discrimination is worsened and the dependence of frequency discrimination upon I.A.D. can be observed at larger differences in the direction of the two tones. The last simple experiment does not provide quantitative data on how the dimensions of the complex tone are perceived, but it does give evidence to Smith's observation [19] with his fishing sonar that two shoals of fish could be detected at the same time and their sizes were comparable.

REFERENCES

[1] CHAPMAN, S. Size, shape and orientation of sonar targets measured remotely. American Journal of Physics, Vol.39, 1971, pp.1181-1189.

[2] ROWELL, D. Auditory display of spatial information. Ph.D. Thesis, University of Canterbury, 1970.

[3] DO, M.A. and KAY, L. Resolution in an artificially generated multiple object auditory space using new auditory sensation. Acustica, Vol.36, 1976/77, pp.9-15.

[4] FLETCHER, H. Auditory patterns. Review of Modern Physics, Vol.12, 1940, pp.47-65.

[5] HAWKINS, J.E. and STEVENS, S.S. The masking of pure tone and speech by white noise. J. Acoust. Soc. Am., Vol.22, 1950, pp.6-13.

[6] HAMILTON, P.M. Noise masked threshold on a function of tonal duration and masking noise bandwidth. J. Acoust. Soc. Am., Vol.29, 1957, pp.506-511.

[7] GREENWOOD, D.D. Auditory masking and the critical band. J. Acous. Soc. Am., Vol.33, 1961, pp.484-502.

[8] SWETS, J.A., GREEN, D.M. and TANNER, W.P. Jr. On the width of critical bands. J. Acous. Soc. Am., Vol.34, 1962, pp.108-113.

[9] DE BOER, E. Note on the critical bandwidth. J. Acous. Soc. Am., Vol.34, 1962, pp.985-986.

[10] LICHTE, W.H. and GRAY, R.F. The influence of overtone structure on the pitch of complex tones. J. Exper. Psychol. Vol.49, 1955, pp.431-436.

[11] LICKLIDER, J.C.R. Auditory frequency analysis. Symposium on information theory. The Royal Institution, London, 1955, London, Butterworths Scientific Publications, 1956.

[12] GOLDSTEIN, J.L. Auditory nonlinearity. J. Acoust. Soc. Am., Vol.41, 1967, pp.676-688.

[13] HALL, J.L. Auditory distortion products $f_2 - f_1$ and $2f_1 - f_2$. J. Acoust. Soc. Am., Vol.51, 1972, pp.1863-1871.

[14] HALL, J.L. Monaural phase effect: cancellation and reinforcement of distortion products $f_2 - f_1$ and $2f_1 - f_2$. J. Acoust. Soc. Am., Vol.51, 1972, pp.1872-1881.

[15] ZWICKER, E., FLATTORP, G. and STEVENS, S.S. Critical bandwidth in loudness summation. J. Acoust. Soc. Am., Vol.29, 1957, pp.548-577.

[16] ZWICKER, E. Subdivision of the audible frequency range into critical bands. J. Acoust. Soc. Am., Vol.33, 1961, p.248.

[17] SCHARF, B. Critical bands and the loudness of complex sounds near threshold. J. Acoust. Soc. Am., Vol.31, 1959, pp.365-390.

[18] SCHARF, B. Loudness of complex sounds as a function of the number of components. J. Acoust. Soc. Am., Vol. 31, 1959, pp.783-785.

[19] SMITH, R.P. Transduction and audible displays for broad band sonar systems. Ph.D. Thesis, University of Canterbury, New Zealand, 1973.

CHAPTER 6

DISCRIMINATION OF COMPLEX TONES IN AN AUDITORY SPACE CONTAMINATED BY NOISE

CHAPTER 6

DISCRIMINATION OF COMPLEX TONES IN AN AUDITORY SPACE CONTAMINATED BY NOISE

6.1 INTRODUCTION

In an attempt to study the auditory discrimination of spatial information when realistic targets of finite sizes were viewed, experimenting on discrimination performance using complex tones was carried out and described in Chapter 6. Hence an approach to more realistic conditions could be continued by inserting noise into the auditory space. It was found that by partially correlating two independent noise sources of limited bandwidths (0 - 5000 Hz), the simulation of a noise-contaminated auditory space could be realised. The auditory sensation produced by this arrangement was discovered to be very similar to that obtained when listening to the reverberation in the binaural CTFM Sonar (volume reverberation in the fishing sonar). The main purpose of the experiment in this chapter is to measure the capability of discrimination of two complex tones simultaneously presented in the noise contaminated auditory space. It will be shown that as the signal-to-noise ratio (S/N measured over the signal frequency band) decreases, the frequency discrimination of a subject is gradually worsened until a certain value of the S/N ratio is reached, below which the discrimination performance declines very fast and the high tone disappears. The discrimination of the frequencies of 1000, 2000 and 3000 Hz is measured and presented as a function of the S/N ratio, from 7 dB to 37 dB in steps of 5 dB. The extrapolation of this function shows

that to obtain the optimal performance (same as in quiet) a very high S/N ratio is required (larger than 80 dB). At low S/N ratio, when the high tone is entirely masked, its unperceivable presence can cause the faint low tone shifted in direction.

6.2 A DIRECTIONAL NOISE FIELD

It was first shown by Kay [1] that the background noise of the CTFM Sonar within a certain bandwidth due to sea reverberation could be assumed equivalent to band limited white noise. Later, Smith [2] discussed the applications of this result to the two kinds of reverberation (volume reverberation and bottom reverberation) and proposed that experiments in auditory detection of signals against white noise would be relevant to sonar detection problems. Figures 1a and b show the typical spectra of noise due to reverberation given by an FM sonar working at deep water and shallow water, respectively.

Since the system being considered is a binaural sonar, the position of each individual scatterer is uniquely informed by the frequency and I.A.D. of an audible signal in the auditory space. When numerous gatherings of scatterers are simultaneously illuminated by the sonic beam of the sonar, the background of noise due to reverberation from these scatterers must be correspondingly confined into a certain area of the subject's auditory space as described in Figure 2. The sensation obtained when listening to (i) this directional noise field, is very different from the sensation obtained when listening to (ii) either a noise source with both ears, (iii) or two independent noise sources, one feeding each ear.

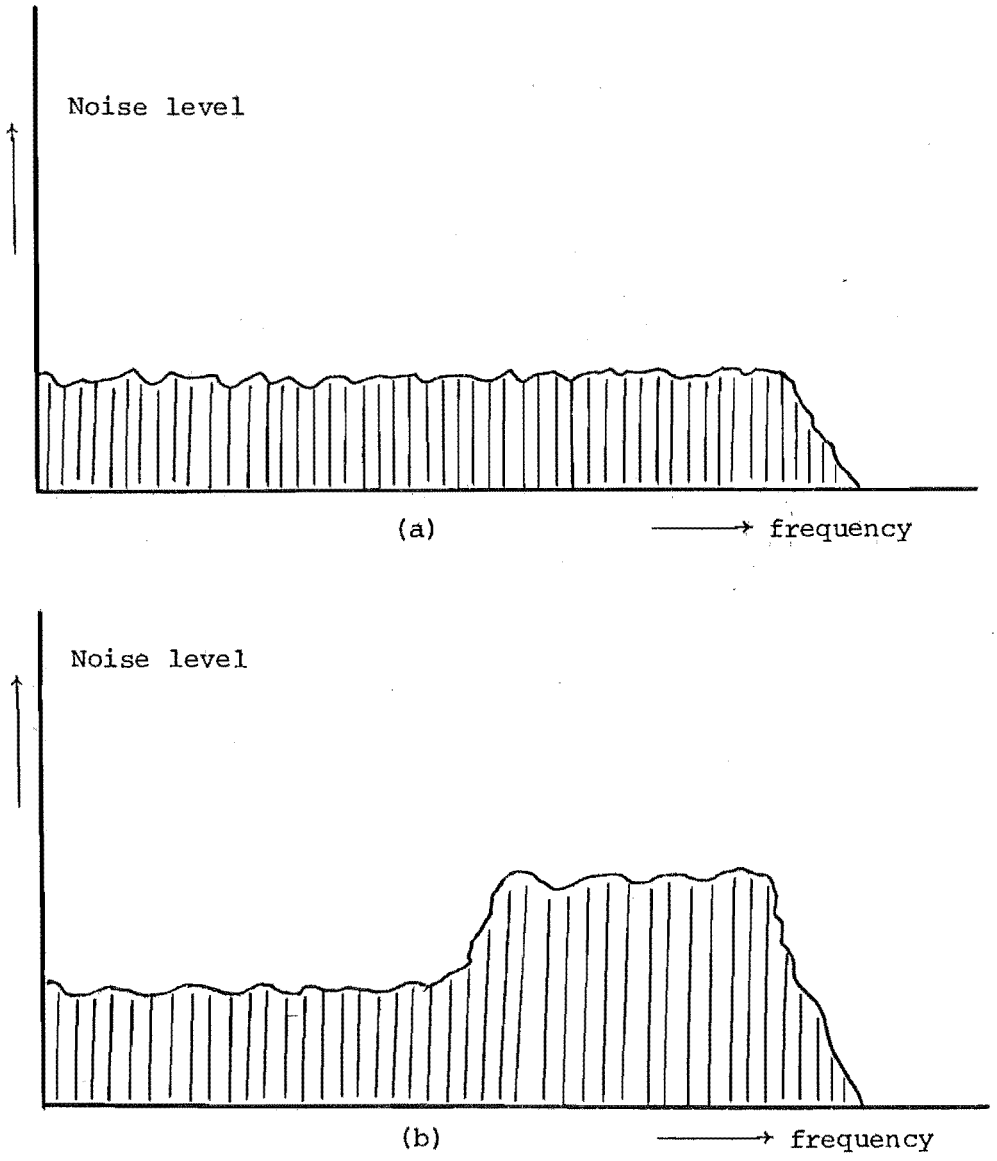


Fig. 1. Spectra of noise due to: (a) Volume reverberation, (b) Volume reverberation, in a CTFM Sonar.

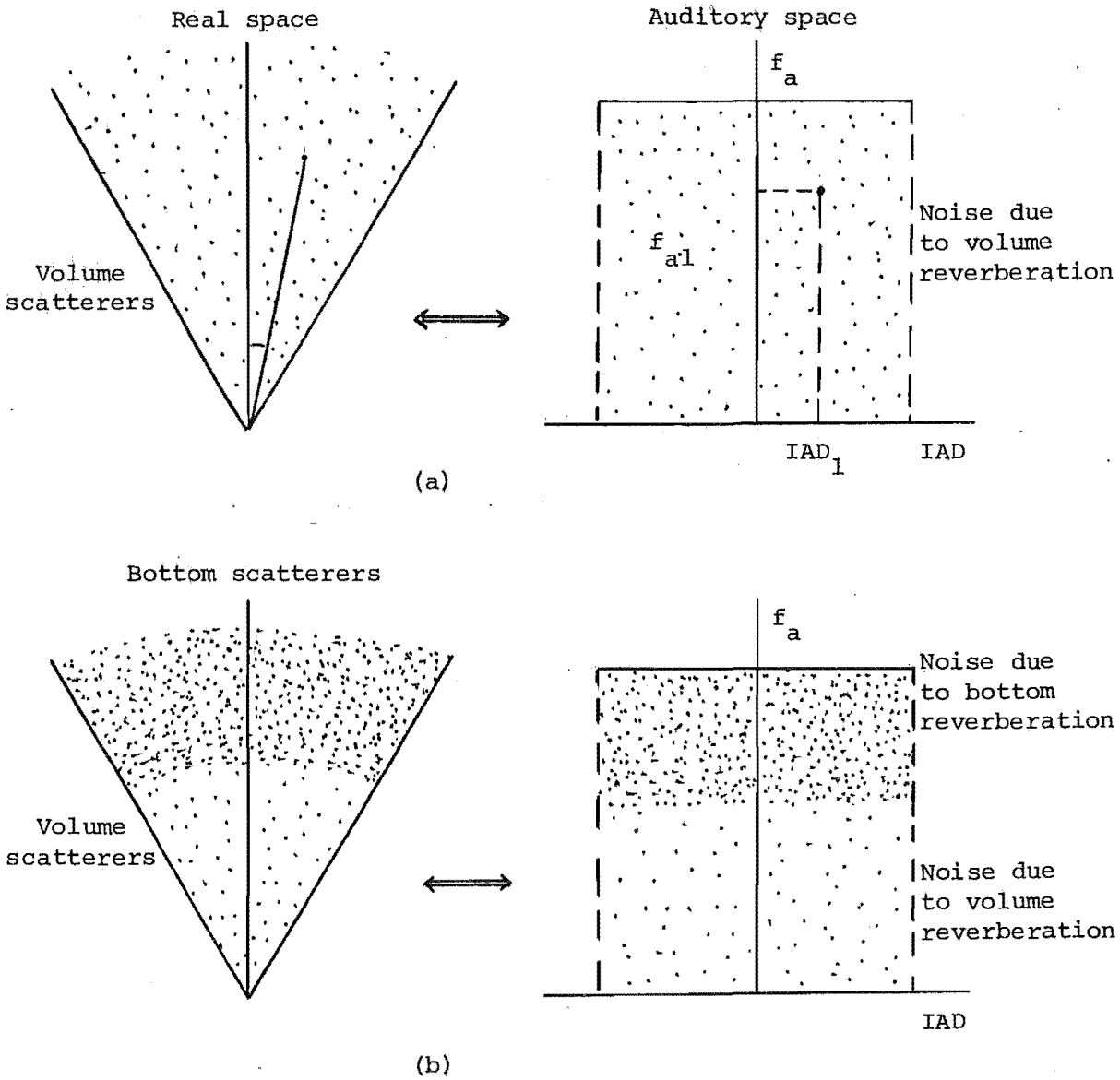


Fig. 2. Transformations from reverberant real spaces to noisy auditory spaces.

In case (i) the subject has an impression that the noise occupies a specific area in front of himself, but he feels that the noise appears either at a single direction only, in case (ii), or omnidirectionally (at all directions), in case (iii). These sensations are similar to the phenomena described by Sayers and Cherry [3,4] as partial fusion (case (i)), complete fusion (case (ii)), and independence (case (iii), i.e. complete lack of fusion).

Simulation of a Directional Noise Field

Since using one noise source can create noise at one direction only, to simulate a directional noise field as described seems to require a large number of noise sources, each set at a specific value of I.A.D. between ± 15 dB. However, when listening to two partially dependent noise sources arranged as in Figure 3, a sensation very similar to that of a directional noise field could be obtained. In Figure 3, $n_1(t)$ and $n_2(t)$ are two independent noise sources of limited bandwidth, say 5000 Hz. The coefficient α of the two Left and Right stimuli, $n_2(t) + \alpha n_1(t)$ and $n_1(t) + \alpha n_2(t)$, determines the degree of fusion of the stimuli. As $\alpha = 0$, there is a complete lack of fusion of the stimuli, and noise is heard in all directions. At $\alpha = 1$, the stimuli are completely correlated, subject hears the noise right in the middle of his field of view. Thus, increasing the value of α from 0 to 1 is equivalent to narrowing down the noise field or increasing the concentration of noise in the forward direction. An explanation for this effect is proposed in Appendix 3. Setting α at the value from 0.18 to 0.25, it was found that noise could be concentrated in a beam of about 60 degrees wide in front of the observer.

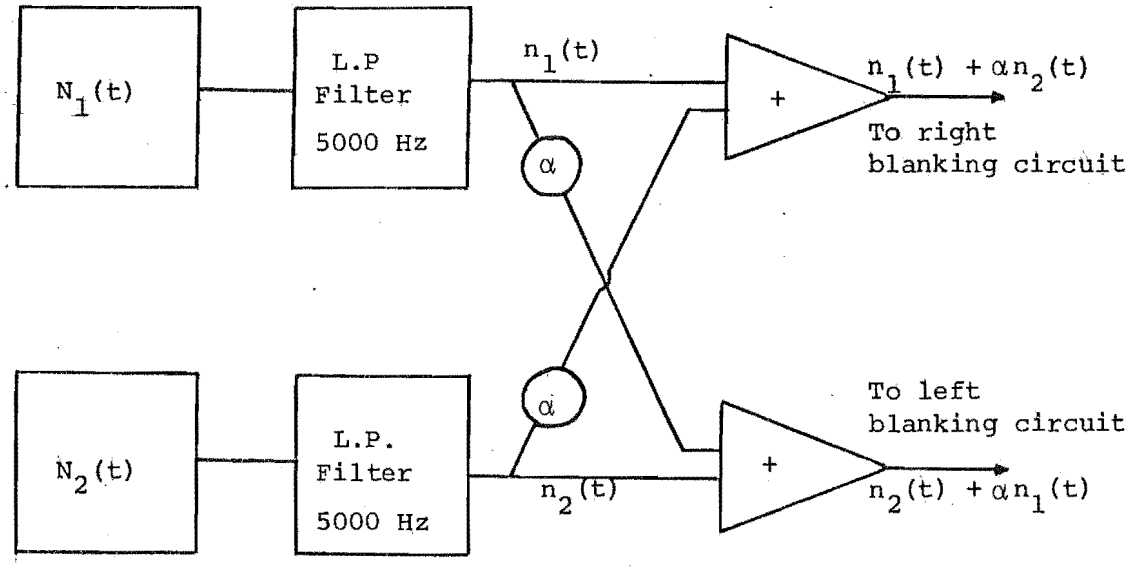


Fig. 3. Simulation of a directional noise field.

(Five subjects listened to this noise field and agreed that the set up produced directional noise more or less as described.)

It should be noted that $20 \log \alpha = -15$ dB when $\alpha = 0.18$. The two sources $n_1(t)$ and $n_2(t)$ instead of being heard as two separate sources at 15 dB to the right and 15 dB to the left respectively were fused together to fill up the gap inside.

6.3 DISCRIMINATION OF TWO COMPLEX TONES IN AN AUDITORY SPACE CONTAMINATED BY NOISE

An experiment was carried out to measure the capability of discrimination of complex tones in the noise-contaminated auditory space at different S/N ratios. Noise used in this experiment was restricted in the frequency band 0 - 5000 Hz, and its level was fixed at 50 dB (each channel). The complex tones were of 100 Hz bandwidths and 2 dB extents in I.A.D. Their levels were dropped in steps of 5 dB from 70 dB to 40 dB after each measurement of frequency discrimination at different values of I.A.D. and frequency was completed.

In the experiments using pure tones and masking noises, the S/N ratio was frequently specified to be the signal to noise ratio per Hz. Thus in this experiment with complex tones, the signal to noise ratio over the signal bandwidth (S/N - S.B) would be the ratio of the signal level and the level of noise contained in a frequency band equal to the signal bandwidth. The S/N - S.B was then reduced from 37 dB to 7 dB in this experiment.

6.3.1 Apparatus and Procedure

The apparatus used to generate two complex tones, the stationary tone g_{ac} and the descending tone $f_{ac}(t)$, were exactly the same as that described in Chapter 6.

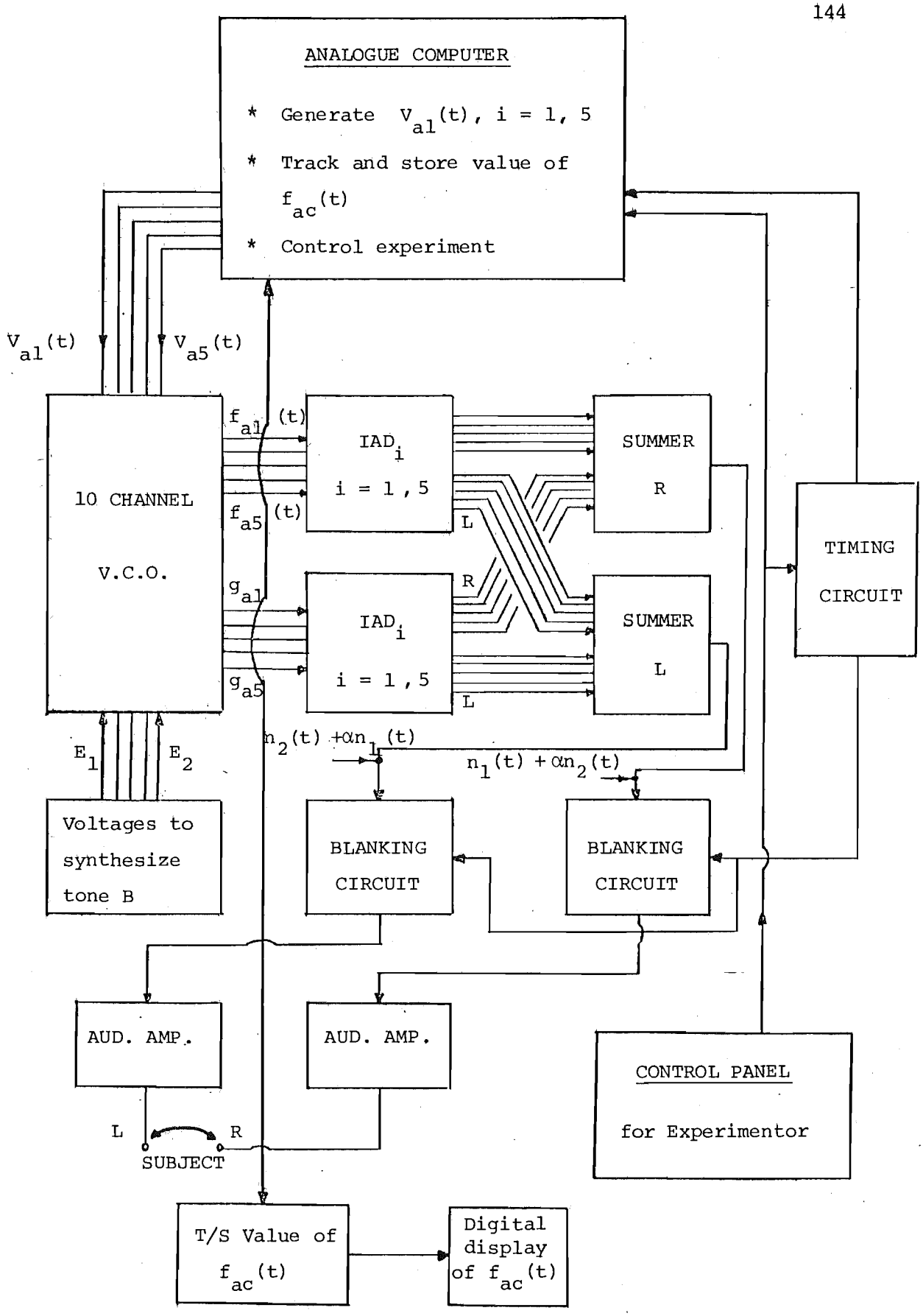


Fig. 4. Arrangement of equipment for the experiment.

The noises, $n_1(t) + \alpha n_2(t)$ and $n_2(t) + \alpha n_1(t)$, used to simulate a directional noise field, and the signals $f_{ac}(t)$ and g_{ac}, were added up at the inputs of the blanking circuits as in Figure 4. The conversion curve of electral power and sound level was measured before the tests were carried out, so during the experiment the sound level could be read on an R.M.S. Voltmeter.

The experimental procedure was based on the phenomenon described by Figure 5. The subject listened to two complex tones, the descending tone $f_{ac}(t)$ and the stationary tone g_{ac}, where $f_{ac}(o) = g_{ac}$, in the noise field, and pressed a switch to store the value of $f_{ac}(t)$ at the instant, t_1, that the descending tone was perceived to slide to the left or right of the initial complex image. The subject was then questioned about the direction of shift of either the low tone $f_{ac}(t)$ or the high tone g_{ac}. The frequency discrimination, $f_{ac} = g_{ac} - f_{ac}(t_1)$, was measured at the frequencies of g_{ac} = 1000, 2000 and 3000 Hz, for three combinations of IAD_c (direction one of the low tone) and JAD_c (direction one of the high tone): (IAD_c, JAD_c) = (0 dB, + 6 dB), (+ 4 dB, - 10 dB) and (+ 8 dB, - 2 dB). Other combinations of IAD's were randomly used so that subject could not memorize any particular binaural patterns while being tested.

6.3.2 Results

Three subjects were used in the experiment, and the average results over six times of recordings for each subject are shown in Tables 1, 2 and 3 of Appendix 4. Their frequency discrimination performances are described as the functions of S/N - S.B in Figures 6, 7 and 8. In each figure, there are three groups of curves corresponding to the subject's performance at 1000, 2000 and 3000 Hz.

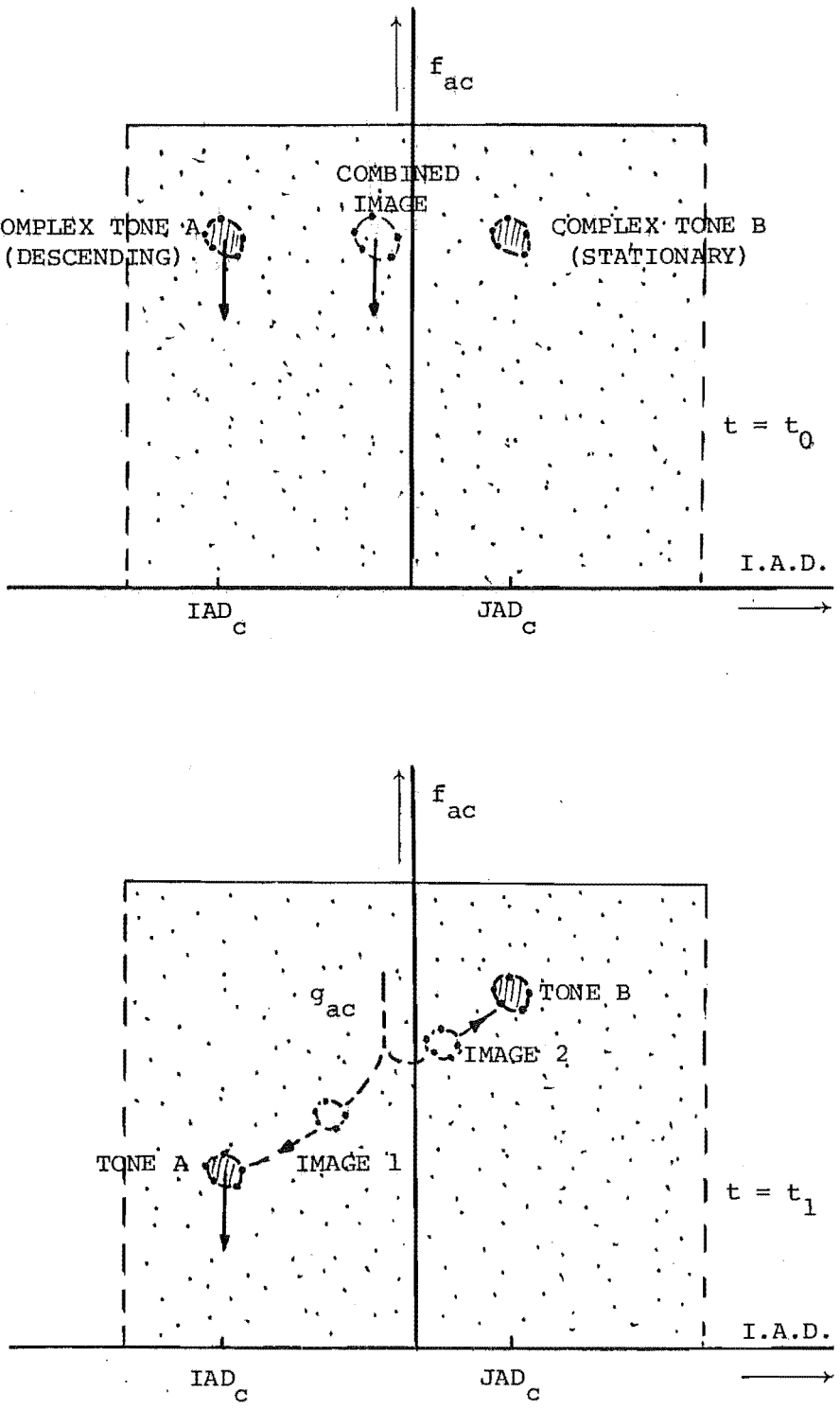


Fig. 5. Discrimination of complex tones in an auditory space contaminated by noise.

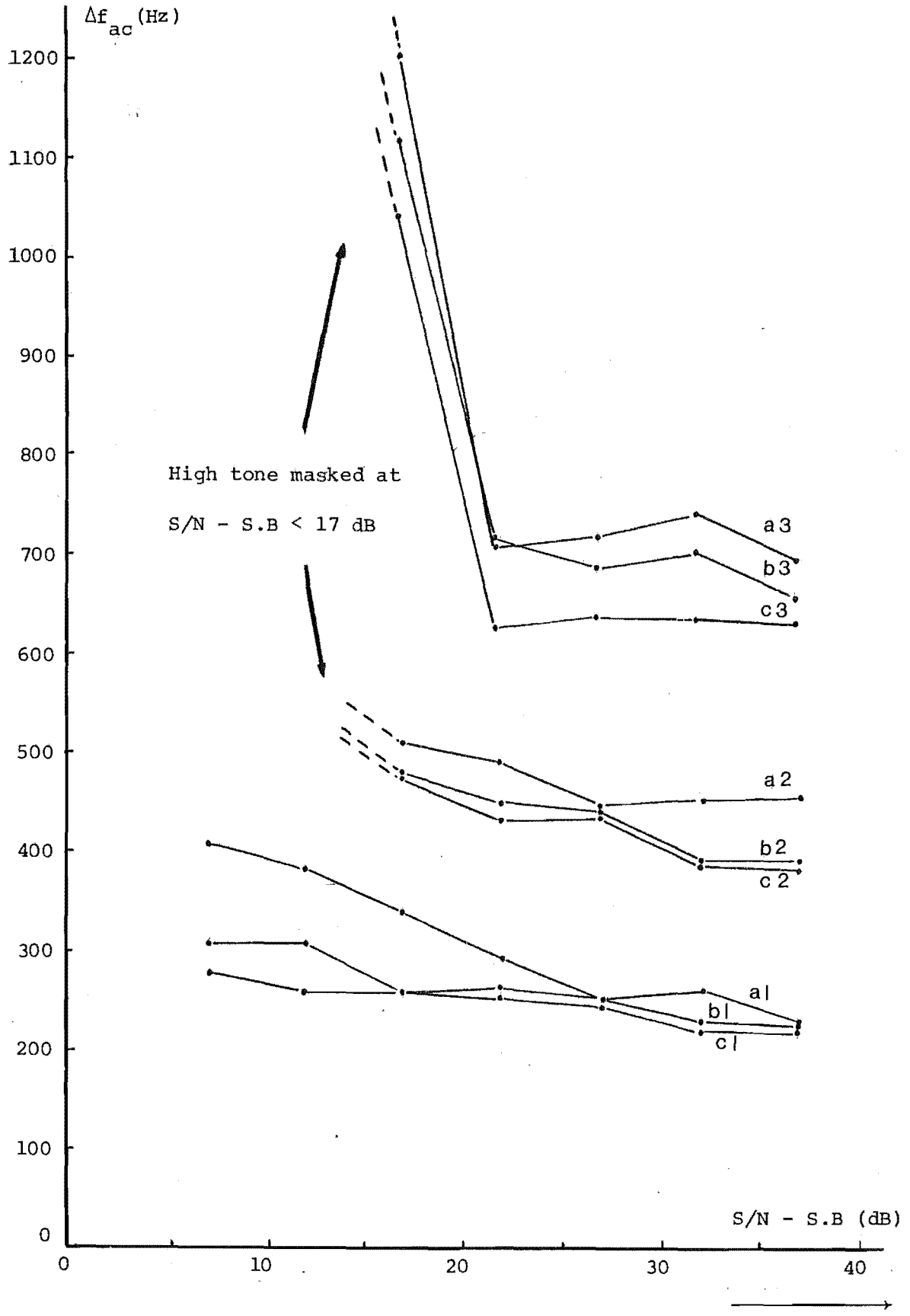


Fig. 6. Discrimination performance of subject S_1 at 1000 Hz (a1, b1, c1), 2000 Hz (a2, b2, c2) and 3000 Hz (a3, b3, c3). (See Table 1.)

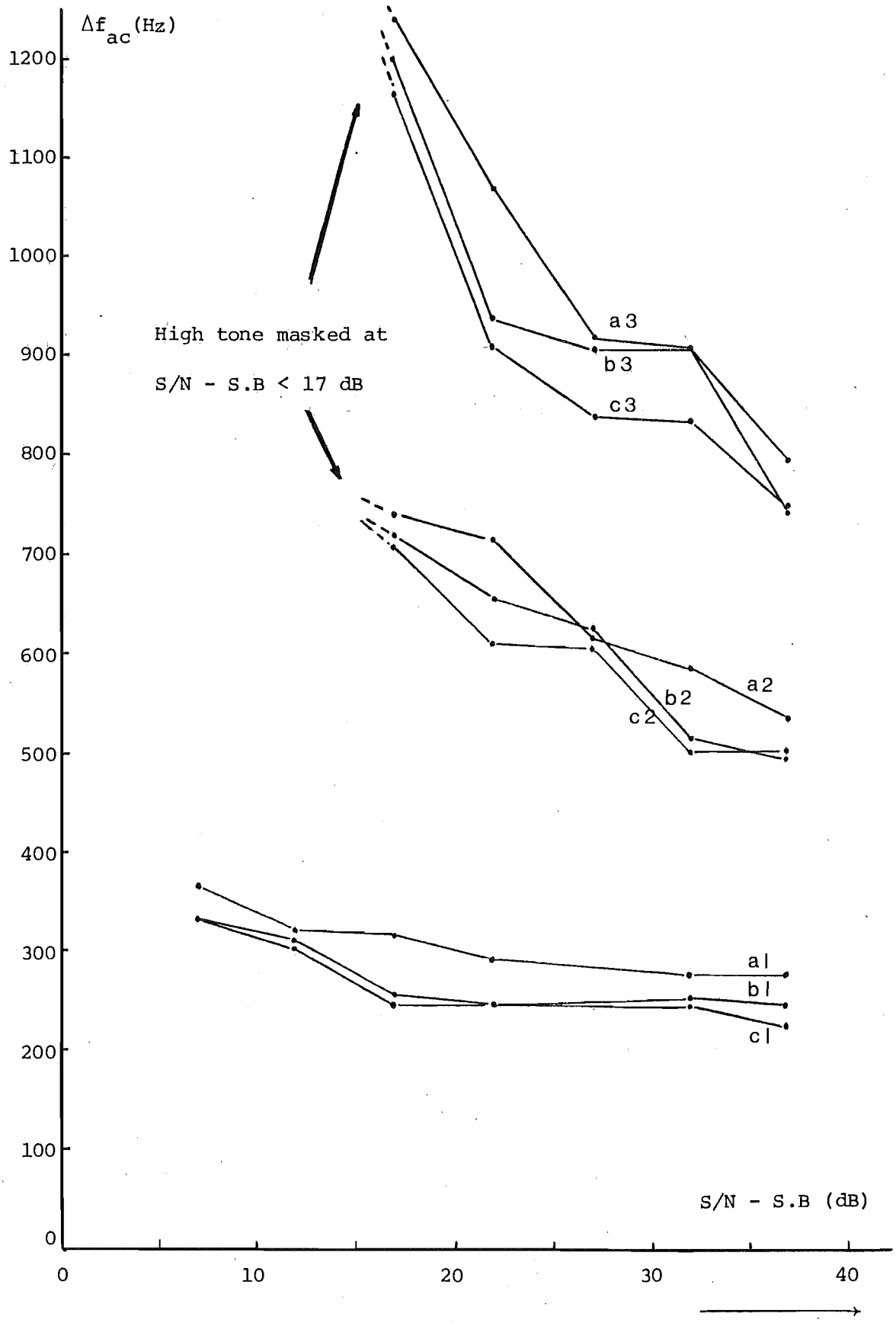


Fig. 7. Discrimination performance of subject S_2 at 1000 Hz (a1, b1, c1), 2000 Hz (a2, b2, c3) and 3000 Hz (a3, b3, c3).

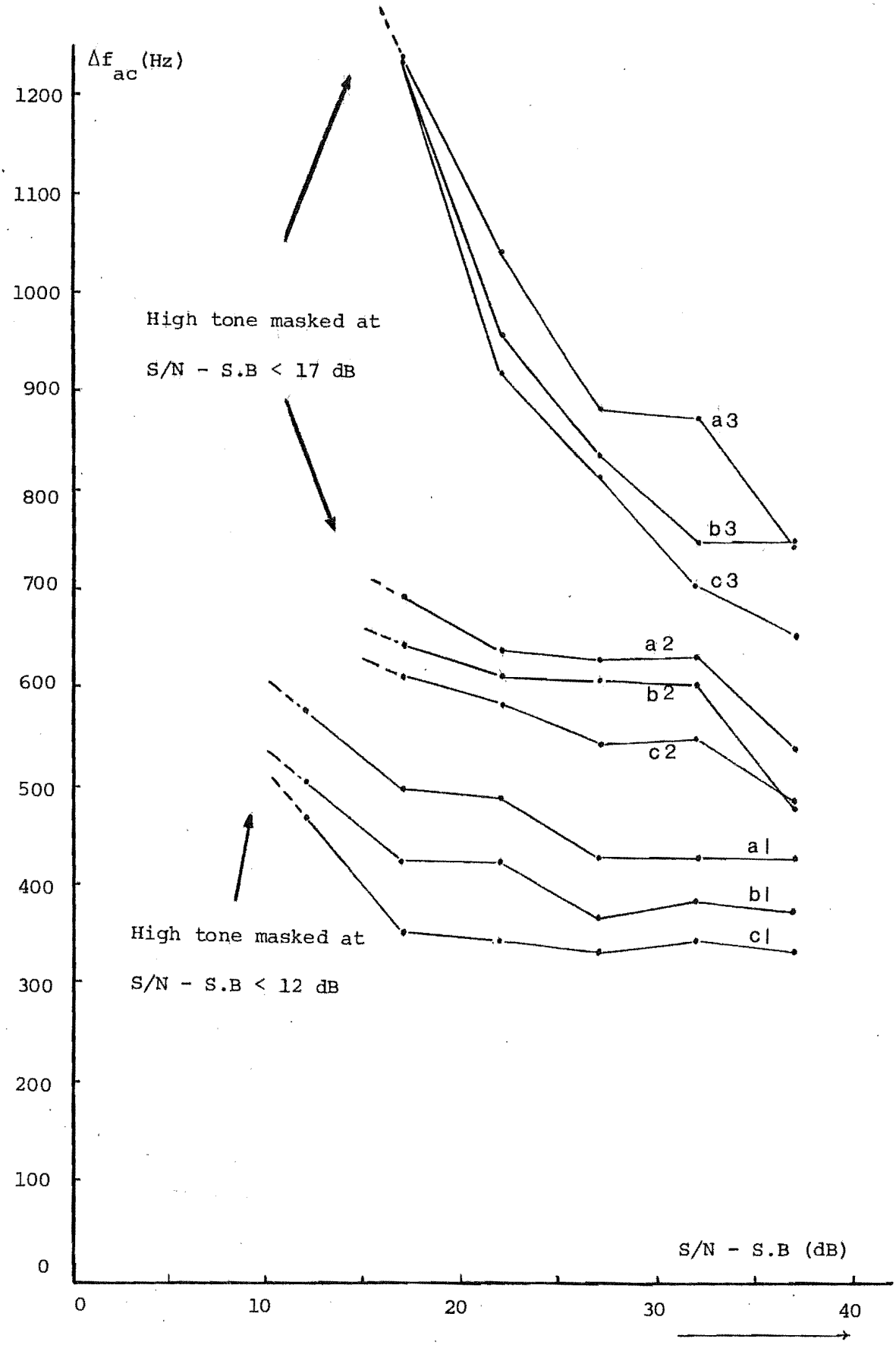


Fig. 8. Discrimination performance of subject S_3 at 1000 Hz (a1, b1, c1), 2000 Hz (a2, b2, c2) and 3000 Hz (a3, b3, c3).

Each group consists of three curves corresponding to three different combinations of IAD's described above. The arrangement of the curves is as follows:

TABLE 1. ARRANGEMENT OF THE FREQUENCY DISCRIMINATION CURVES

g_{ac} Hz	IAD_c dB	JAD_c dB	frequency discrimination curve
1000	0	6	a 1
"	8	- 2	b 1
"	4	- 10	c 1
2000	0	6	a 2
"	8	- 2	b 2
"	4	- 10	c 2
3000	0	6	a 3
"	8	- 2	b 3
"	4	- 10	c 3

One of the results found in the previous experiments of discrimination of complex tones under noise-free conditions which can be seen here is that Δf_{ac} slightly increases as the difference in I.A.D. of two tones is reduced. In general, as the S/N - S.B decreases, the capability of discrimination is gradually worsened (i.e. Δf_{ac} increases) until a certain value of S/N - S.B is reached, the capability of discrimination declines sharply, and the high tone is entirely masked. The detailed observation is as follows:

- At g_{ac} = 3000 Hz , the high tone was entirely masked as the S/N - S.B was reduced to 17 dB. Subject pressed the button to record the value of $f_{ac}(t_1)$ when the low tone was heard to raise clearly and slide to a determinable direction.

- at g_{ac} = 2000 Hz , subject reported the disappearance of the high tone in about 50% of the number of tests, when the S/N - S.B was reduced to 17 dB.

- at g_{ac} = 1000 Hz , the experiment was carried out until the S/N - S.B was reduced to 7 dB. At this signal to noise ratio, subjects 1 and 2 could still perceive both tones and discriminate them easily, but subject 3 lost the high tone in about 50% of the number of tests when the S/N - S.B was equal to 12 dB.

6.4 DISCUSSION

Subjects S_1 and S_2 used in this experiment are the same subjects in previous experiments on the discrimination of complex tones in quiet conditions. Hence their performances in quiet and in noise are going to be compared. The results obtained from subjects S_1 and S_2 at different combinations of IAD_c and JAD_c are now averaged, and the idealized curves of discrimination performance are plotted as the functions of S/N - S.B (see Figure 9). In quiet conditions, the average discrimination performances of these two subjects over the similar combinations of IAD's are Δf_{ac} = 115, 165 and 210 Hz at g_{ac} = 1000, 2000 and 3000 Hz respectively. The extrapolation of the curves shows that it may require an S/N - S.B as high as 80 dB in order that these subjects can perform their best discriminations. Such a requirement has not been achieved in any wide beam FM sonar with targets such as fish shoals (see Chapters 8 and 9).

As the S/N - S.B is reduced, there would be a certain limit where the high tone is entirely masked. This limit as shown in the experiment result is obviously higher than the masked threshold of a tone when it is presented with noise only.

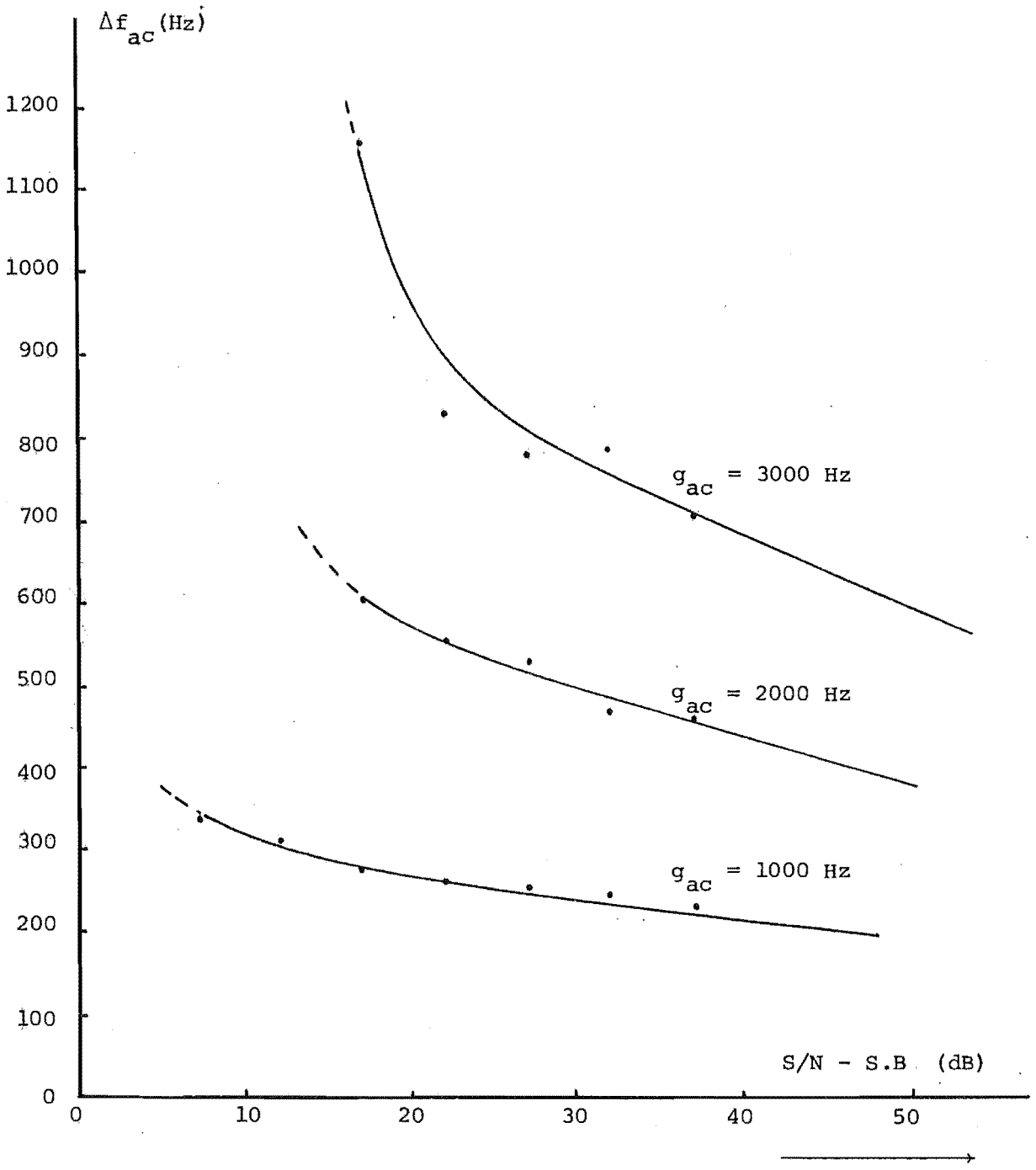


Fig. 9. Average Δf_{ac} of S_1 and S_2 over different values of IAD's.

Indeed, using the definition of a critical band as the bandwidth of noise whose overall energy is equal to the energy of a tone in the center of the band when the tone is just barely masked by the noise [5], and accepting the hypothesis proposed by Schafer and Gales, Marill, and Green [6,7,8] that the power of a band limited signal is summated if its bandwidth is smaller than a C.B, then the masked threshold of a complex tone must be:

$$S/N - S.B_{\text{at threshold}} = 10 \log \left(\frac{\text{critical bandwidth}}{\text{signal bandwidth}} \right)$$

Using the results given by Zwicker et al. [9,10] and Scharf [11,12] that the critical bandwidths of loudness summation at 1000, 2000 and 3000 Hz are respectively equal to 160, 300 and 470 Hz, the masked thresholds of the complex tones of 100 Hz bandwidths must respectively be 2, 5 and 7 dB, at 1000, 2000 and 3000 Hz. The shift of the masked threshold of the high tone is perhaps due to the presence of the low tone, even though two tones are well separated in frequency. Wegel and Lane [13], when sounding two pure tones simultaneously, found a similar effect that the hearing threshold of the high tone in quiet was shifted by the presence of the low tone.

During the course of the experiments it was also found that at low S/N - S.B, when the high tone is entirely masked, the direction of the low tone is shifted in the auditory space. Indeed, for example, at the beginning of a trial, tone A (descending) of 0 dB IAD and tone B (stationary) of + 10 dB (Right) are of the same frequencies; a combined image would be perceived by the subject somewhere in the right hand. As tone A descends, beating between two tones is gradually reduced until a certain frequency difference is reached and the subject hears the low tone slide definitely to the left hand instead of the centre of the auditory space, while the

high tone is not perceivable. It is obvious that the presence of the high tone at one side of the auditory space causes the masking background noise to be unbalanced. The different reductions in the perceived amplitudes of the low tone in two ears by masking, results in a change of its perceived direction. A similar effect was found by Sanders [14].

6.5 CONCLUSION

Throughout the experiment it can be concluded that

(i) Within the range of the S/N - S.B from 20 dB to 40 dB the frequency discrimination of the complex tones is about two to four times worse than that in quiet conditions. This performance is gradually improved with increasing the S/N - S.B. However, a request of an 80 dB S/N - S.B for the optimal performance is perhaps obtainable only in the blind aid, not in the fishing sonar.

(ii) The frequency discrimination is gradually worsened as the S/N - S.B decreases till a certain value below which the performance declines very fast and the high tone disappears.

(iii) The high tone is completely masked at an S/N - S.B of 10 dB higher than its masked threshold when it is presented alone, and its unperceivable presence can cause the faint low tone to be shifted in direction.

Obviously, noise affects the discrimination performance significantly. Although this experimental result is less relevant to the case of the blind aid, where the reverberation from wall, bush and fence provides necessary information for the sonic-glasses user and the environmental noise is entirely different, it does apply to the cases of the child-aid (an adaptation of the sonic glasses for the blind child) and the fishing sonar.

The simulation can resemble either a child sitting on the floor and picking up toys from the carpet, or a boat tracking one of the fish shoals by using CTFM sonar.

Chapter 7 will investigate the problem of target detection in an underwater wide beam FM sonar, and Chapter 8 will compare the masked thresholds of complex tones in static and dynamic cases.

REFERENCES

[1] KAY, L. An experimental comparison between pulse and frequency modulation echo ranging systems. J. Brit. I.R.E., Vol.19, 1960, pp.785-796.

[2] SMITH, R.P. Transduction and audible displays for broad band sonar systems. Ph.D. Thesis, University of Canterbury, New Zealand, 1972.

[3] CHERRY, E.C. and SAYERS, B. McA. Human 'cross-correlator' - A technique for measuring certain parameters of speech perception. J. Acoust. Soc. Am., Vol.28, 1956, pp.889-895.

[4] SAYERS, B. McA. and CHERRY, E.C. Mechanism of Binaural fusion in the hearing of speech. J. Acoust. Soc. Am., Vol.29, 1957, pp.973-987.

[5] FLETCHER, H. and MUNSON, W.A. Relation between loudness and masking. J. Acoust. Soc. Am., Vol.9, 1937, pp.1-10.

[6] SCHAFER, T.H. and GALES, R.S. Auditory masking of multiple tones by random noise. J. Acous. Soc. Am., Vol.21, 1949, pp.392-399.

[7] MARILL, T. Detection theory and psychophysics. Tech. Rept. No. 319, Research Laboratory of Electronics, M.I.T. Cambridge, Massachusetts, 1956.

[8] GREEN, D.M. Detection of multiple component signals in noise. J. Acoust. Soc. Am., Vol.30, 1958, pp.904-911.

[9] ZWICKER, E., FLOTTORP, G. and STEVENS, G.G. Critical bandwidth in loudness summation. J. Acoust. Soc. Am., Vol.29, 1957, pp.548-557.

[10] ZWICKER, E. Subdivision of the audible frequency range into critical bands (frequenzgruppen). J. Acoust. Soc. Am., Vol.33, 1961, p.248.

[11] SCHARF, B. Critical bands and the loudness of complex sounds near threshold. J. Acoust. Soc. Am., Vol.31, 1959, pp.365-370.

[12] SCHARF, B. Loudness of complex sounds as a junction of number of components. J. Acoust. Soc. Am., Vol.31, 1959, pp.783-785.

[13] WEGEL, R.L. and LANE, C.F. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev., Vol.23, 1924, pp.266-285.

[14] SANDERS, S.D. The distortion of an auditory display with noise. M.E. Thesis, University of Canterbury, New Zealand, 1974.

CHAPTER 7

BASIC CONSIDERATIONS ON DETECTION PROBLEMS IN WIDE BEAM CTFM SONAR

CHAPTER 7

BASIC CONSIDERATIONS ON DETECTION PROBLEMS IN WIDE BEAM CTFM SONAR

7.1 INTRODUCTION

In sonar design processes, it is frequently necessary to estimate at the position of the sound source the intensities of the sonic waves reflected from targets, scatterers and boundaries. Terminologies such as back scattering cross-section, target strength and reverberation level have become familiar to sonar designers as the indications of strength of the target echo (the first two terms and the background (the last term). The designer's objective is to find means for increasing the overall response of the sonar system to the target-echo (also called the signal) and for decreasing the response of the system to the background (or noise) - in other words to increase the signal to noise ratio.

A sonar system can be designed to serve specific purposes such as detection, discrimination of multiple targets, classification (i.e., determining the character of a target); or more particularly, fish finding. Depending on the desired purpose, a certain signal to noise ratio is required so that the function can be performed. For example, in Chapter 6, it was shown that in a CTFM Sonar with binaural display, the discrimination of two targets could be accomplished only when the signal to noise ratio would be about 10 dB higher than that required for detection. The condition for the detection to be just accomplished can be described by:

Signal level = background "masking" level

where the term "masking" implies that not all the background interferes with the signal, but only a portion of it. This portion depends upon the method by which the signal and noise are processed and displayed. For instance, with a pulsed sonar, the extension in range, Δr , of the annular region, which determines the amount of reverberation effectively masking the echo, is such that the scattering produced by all portions of the annular region returns to the source at the same instant of time as the echo. Accordingly, Δr is equal to $C\tau/2$ where C is the velocity of sound and τ is the pulse length [1]. In CTFM Sonar, however, the scattering produced by all portions of the sonic beam from the zero range to the furthest detectable range is simultaneously displayed on a frequency scale. The frequencies of the displayed components are arranged to be correspondingly proportional to the distances of the portions of scatterers. Thus, the signal is effectively masked by a band of noise whose bandwidth is equal to the bandwidth of the signal, provided that the filter is idealised. In other words, it is the extension in range, Δr , of the target that determines the annular region of scatterers from which the reverberation masks the echo effectively.

The main purpose of this chapter is to study the use of CTFM Sonar in fish finding. The first part provides methods for estimating the target strength of a shoal of fish. Noise due to sea reverberation is investigated in the second part. The theoretical prediction of the system performance is then compared to the results of the experiment conducted by Kay et al.* recently.

---

*The experiments have been carried out by Kay et al. of the Electrical Engineering Department, University of Canterbury, New Zealand, since 1975.

7.2 MEASUREMENT OF TARGET STRENGTH AND BACK SCATTERING CROSS-SECTION OF FISH

The parameter target strength referring to the echo returned by an underwater object is defined as:

$$T.S. \equiv 10 \log \frac{I_{rl}}{I_i} \qquad (1)$$

where I_{rl} is the intensity of the sound returned by the target at a distance of 1 metre from its "acoustic center" in some direction, and I_i is the incident intensity from a distance source. According to this definition, I_{rl} must be measured in the direction of the receiver of the sonar system. Hence, target strength is a function of both the incident direction and the direction of the receiver. However the study in this chapter is restricted to the case of closely adjacent source and receiver, thus target strength depends on the incident direction only [1].

An alternative parameter frequently used to measure the echo returned from an object is the "Back scattering cross section". This is defined by Kerr [2] as "The area intercepting that amount of power which, when scattered isotropically, produces an echo equal to that observed from the target". The relationship between the back scattering cross section and the target strength is

$$T.S = 10 \log \frac{\sigma}{4\pi} \qquad (2)$$

where σ is the back scattering cross section, which, in general, differs from the geometrical cross section of a target.

Appendix 5 provides a detailed study of these two parameters and gives the formulas of target strength and back scattering cross section of a number of geometric shapes.

The lengths of some of the better known fish which are caught commercially are given in Table 1. These fish are similar in shape. Each has a swim bladder, which is a gas filled, cigar-shaped envelope situated in the arch of bones of the backbone.

TABLE 1. Lengths of fish [3,4]

Name	Length L (cm)	Name	Length L (cm)
Tuna	70 - 200	Whiting	20 - 40
Cod	45 - 120	Herring	15 - 30
Hake	30 - 120	Pilchard	12 - 25
Haddock	30 - 60	Sprat	8 - 15

Using the scale model technique, Haslett [3,4] measured the back scattering cross-section (also called acoustic cross-section) of fish at 1.48 MHz, 625 kHz and 360 kHz, over a wide range of fish length (i.e. length of fish) (8 - 60 times the wave length λ). His results can be summarised as follows:

At constant frequency, when the length of the fish is increased (or alternatively, for a constant fish-length, as the frequency is raised), the fleshy body of the fish which is initially in the Rayleigh Scattering Region (σ is proportional to $1/\lambda^4$) passes into the geometrical region before the swim bladder and the backbone ($10\ \lambda \leq L \leq 24\ \lambda$). However, for larger fish (or at higher frequency), $24\ \lambda \leq L \leq 60\ \lambda$, the scattering signals from the cylindrical bladder and backbone will exceed that from the body, and σ is proportional to $1/\lambda$. If the ratio L/λ is increased further and a plane reflecting area is involved, the acoustic cross-section is proportional to $1/\lambda^2$. Figure 1 shows the acoustic cross-sections of fish (sticklebacks and guppies) in main aspects (head, tail, dorsal, ventral, etc.) measured at 1.48 MHz.

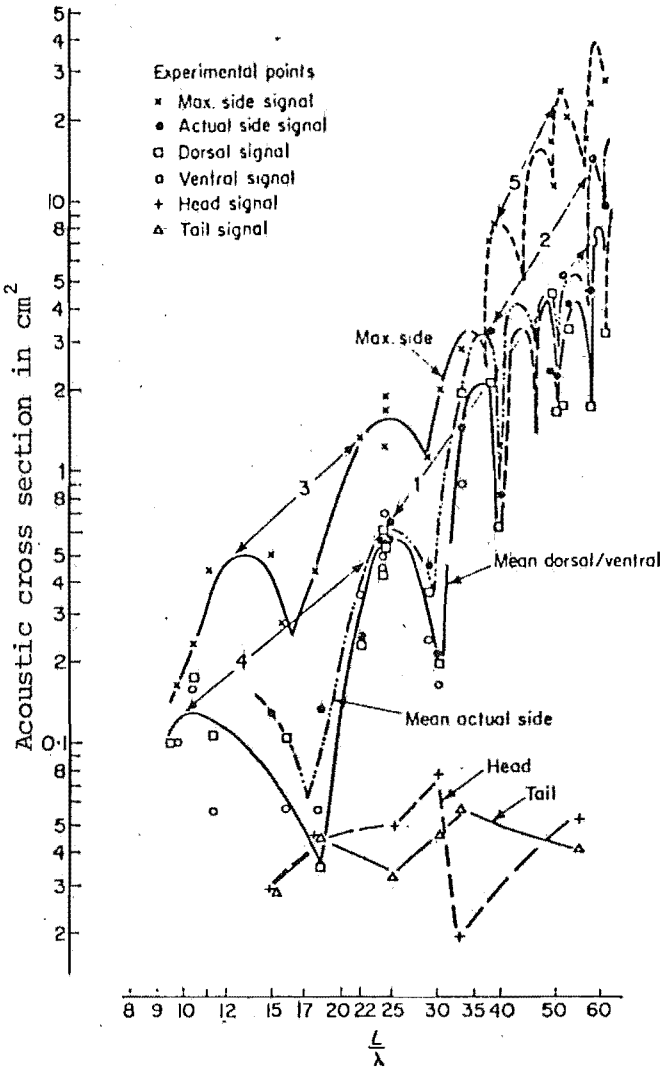


Fig. 1. Acoustic cross-sections of sticklebacks and guppies in main aspects, measured at 1·48 MHz. (L = overall length of fish, λ = wavelength in water.) The numbered straight lines show the general trends of the cross sections at various points:

Line	Slope
1 (dorsal)	2·9
2 (actual side)	3·1
3 (maximum side)	1·8
4 (dorsal)	1·6
5 (maximum side)	4·1

(Reproduced from [3,4])

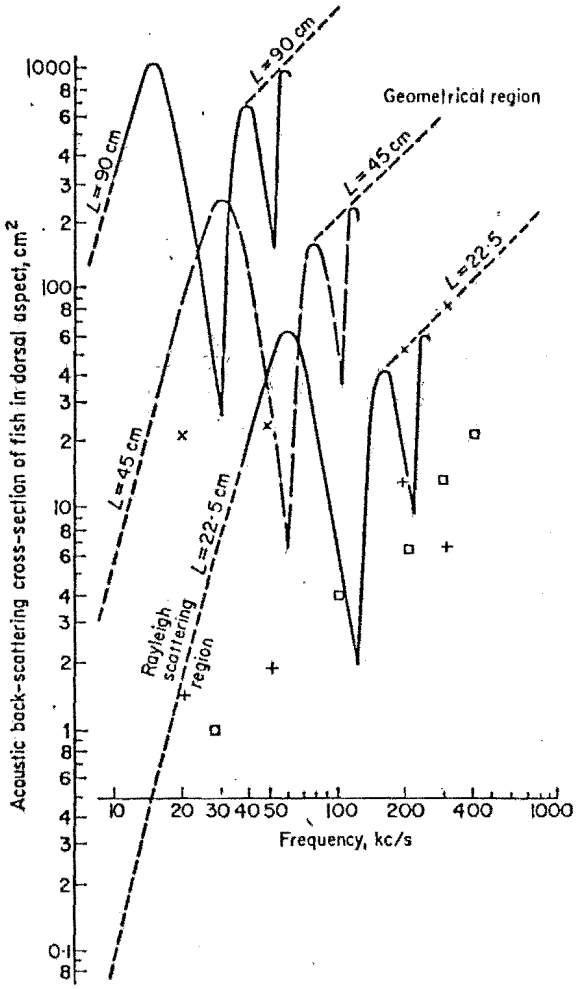


Fig. 2. Approximate frequency responses of fish of lengths 90 cm, 45 cm and 22·5 cm determined by scale-model measurements in the geometrical region. At high frequencies there are many maxima, so the general trend of the maxima is shown as a dotted line on each graph.

These results are compared with those published by Hashimoto and Maniwa for a fish of length 15 cm (□) and by Shishkova for fish of lengths 45 cm (×) and 22·5 cm (+) [5, 9]

(Reproduced from [3,4])

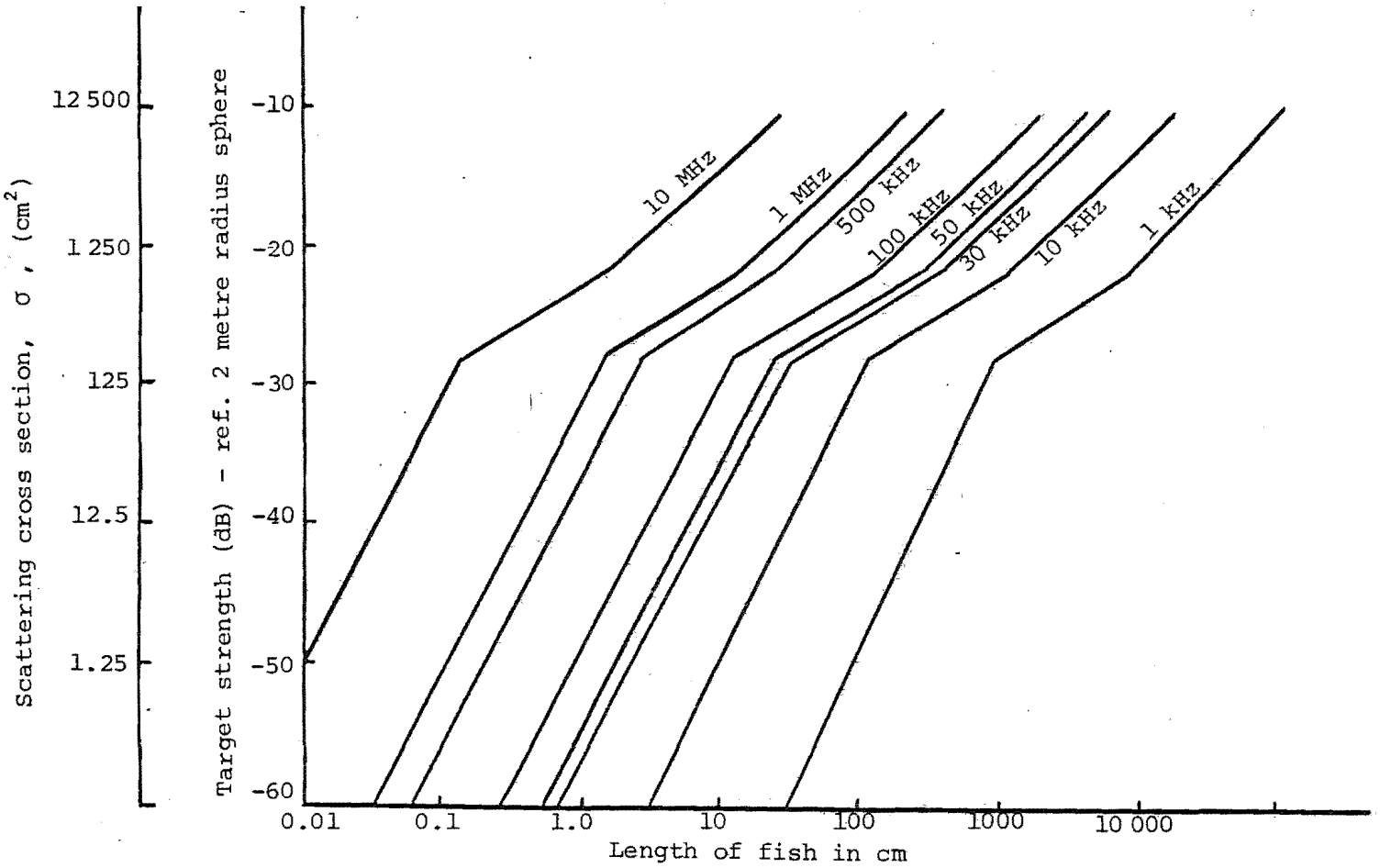


Fig. 3. The relation between scattering cross section and the length of the fish at different frequencies. (Reproduced from [6])

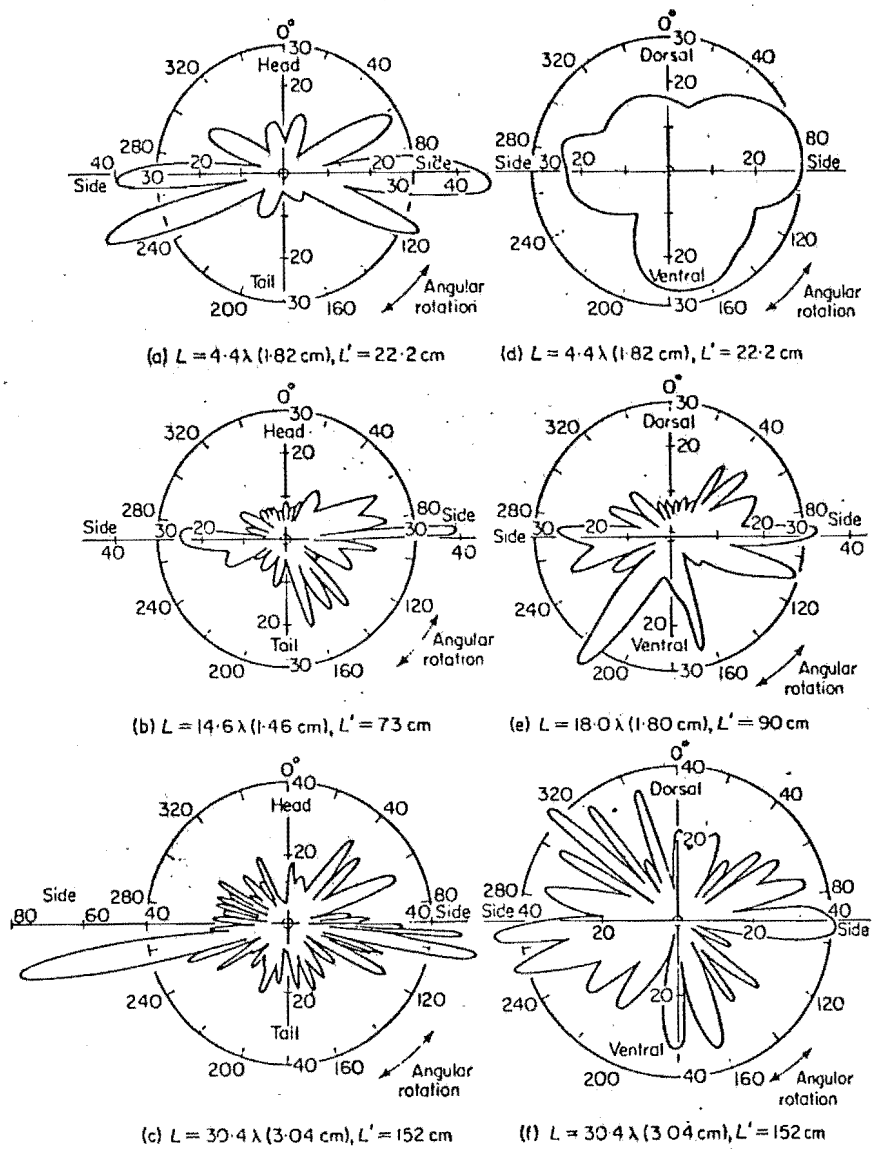


Fig. 4. Back-scattering polar diagrams of fish (sticklebacks) . The readings were taken under the following conditions:

Plane of observation: (a), (b) and (c), horizontal plane; (d), (e) and (f), vertical plane (perpendicular to longitudinal axis of fish).

Frequency: (a) and (d), 360 kHz; (b), (c), (e) and (f), 1·48 MHz.

Radial scale: amplitude (mV): 20 mV in (a) and (d) is equivalent to an acoustic cross-section of $1{\cdot}5 \times 10^{-2}$ cm² at 360 kHz (or $8{\cdot}8 \times 10^{-4}$ cm² at 1·48 MHz); 20 mV in (b), (c), (e) and (f) is equivalent to an acoustic cross-section of $1{\cdot}73 \times 10^{-1}$ cm² at 1·48 MHz.

L = actual length of fish, L' = equivalent full-size length of fish at 30 kHz, λ = wavelength in water

(Reproduced from [3,4])

The frequency response of a fish of any arbitrary length may be calculated from these results using the correct scaling factor for each point. Figure 2 shows the frequency response of fish in dorsal aspect. The lengths of fish are 90 cm, 45 cm and 22.5 cm.

Cushing et al. [6] measured the scattering cross-section of fish at dorsal aspect, and found the results agreeing with Haslett's measurements. Figure 3 shows that each curve consists of three distinguishable parts: (i) $L < 8\lambda$, σ is proportional to $1/\lambda^4$; (ii) $8\lambda < L < 100\lambda$, σ is proportional to $1/\lambda$; and (iii) $L > 100\lambda$, σ is proportional to $1/\lambda^2$. The line joining the peaks of Haslett's curve in the transient part ($8\lambda < L < 24\lambda$) is reasonably fitted to the relation σ oc $1/\lambda$.

Obviously the above results are applicable only to some specific conditions of the sonar system where the aspect of the fish illuminated by the acoustic signal is known. For example, under normal echo-sounding conditions, the dorsal surface of the fish is presented to the acoustic signal. Such conditions are not found in a forward looking sonar, where fish in a shoal are illuminated by the sonic wave under various incident angles, and their target strengths vary during the time they swim or the boat moves (Figure 4 shows the back scattering polar diagrams of fish observed in different planes). Hence, to estimate the target strength of a fish shoal, it is necessary to determine the average target strength of fish over various incident angles of the sonic wave. This is done in the next section.

7.3 DETERMINATION OF THE AVERAGE TARGET STRENGTH AND BACK SCATTERING CROSS-SECTION OF FISH WITH RESPECT TO INCIDENT ANGLE

The essential parts of a fish which contribute most to the echo signal can approximate to the following simple forms:

Body - The body of the fish (less fins) approximates to an ellipsoid of dimensions: $L_1 = 0.93\ L$, $H_1 = 0.195\ L$, and $B_1 = 0.112\ L$; where L is the total length of the fish. The reflectivity of fish flesh is $\beta = 4.4\%$ in fresh water, and 1.9% in salt water (see Appendix 5).

Swimbladder - The swimbladder can approximate to a cylinder of length 0.24 L, and radius 0.0245 L. Its reflectivity is -100% .

Vertebral Column - The core of the vertebral column is a cylinder of length about 0.65 L having a mean diameter of about 0.012 L . The reflectivity of wet fish bone has been found to be 26% (Appendix 5). The signals reflected from the bony spines of the backbone are much smaller than from the core.

Applying the formulae given in Appendix 5, the average back scattering cross-sections of these parts of the fish, and of the whole fish, can be determined as follows:

(i) Very small fish

The back scattering cross-section of a rigid body having its dimensions in Rayleigh scattering region is [7]

$$\sigma = \frac{16\ \pi^3\ V^2}{\lambda^4} \qquad (3)$$

According to Mentzer [10], a cylinder is considered as in the Rayleigh region when its radius is smaller than $\lambda/2\pi$. The radius of the swimbladder is larger than that of the backbone and equal to 0.0245 L.

Thus equation (3) can be used to calculate the scattering cross sections of both the swimbladder and the backbone, provided the fish length L is smaller than $6.5\ \lambda$.

Using the given approximations to parts of a fish, the volume of the swimbladder is:

$$V_1 = 4.53 \times 10^{-4}\ L^3 \tag{4}$$

So its scattering cross section is

$$\sigma_1 = 1.02 \times 10^{-4}\ \frac{L^6}{\lambda^4} \tag{5}$$

Similarly, the volume of the backbone is

$$V_2 = 7.35 \times 10^{-4}\ L^3 \tag{6}$$

Substituting V_2 into equation (3) and multiplying the result with the square of reflectivity of wet bone

$$\sigma_2 = 1.81 \times 10^{-7}\ \frac{L^6}{\lambda^4} \tag{7}$$

At the fish lengths close to $6.5\ \lambda$, the body of the fish is, in fact, in the transient region. However, since it is still closer to the Rayleigh region than to the geometrical region, equation (3) would be employed to calculate its scattering cross section [5]. The volume of the fleshy part of the fish is

$$V_3 = 7.80 \times 10^{-3}\ L^3 \tag{8}$$

So its scattering cross section is

$$\sigma_3 = 5.8 \times 10^{-5}\ \frac{L^6}{\lambda^4} \quad \text{in fresh water} \tag{9}$$

$$\sigma_3 = 1.1 \times 10^{-5}\ \frac{L^6}{\lambda^4} \quad \text{in salt water} \tag{10}$$

Thus the back scattering cross section of the whole fish is

since, in geometrical region, $L > 24\ \lambda$

$$k\ l_1 = \frac{2\pi}{\lambda} \times 0.24\ L > 11.52\ \pi >> \pi$$

The function $\left(\frac{\sin u}{u}\right)^2$ tends to zero very fast, therefore

$$\int_0^{k\ l_1} \left(\frac{\sin u}{u}\right)^2 du \doteq \int_0^{\infty} \left(\frac{\sin u}{u}\right)^2 du = \frac{\pi}{2}$$

Thus

$$\sigma_{SB} = a_1\ l_1 \qquad (17)$$

Similarly, the average scattering cross section of the backbone is

$$\overline{\sigma_{BB}} = a_2\ l_2\ \beta^2 \qquad (18)$$

where a_2 and l_2 are respectively the radius and length of the backbone, $\beta = 26\%$ is the reflectivity of wet bone. Substituting the approximated values of a_1 , l_1 , a_2 , l_2 in equations (17) and (18):

$$\overline{\sigma_{SB}} = 5.88 \times 10^{-3}\ L^2 \qquad (19)$$

$$\overline{\sigma_{BB}} = 2.64 \times 10^{-4}\ L^2 \qquad (20)$$

The average scattering cross section of the whole fish is

$$\overline{\sigma} = 6.14 \times 10^{-3}\ L^2 \qquad (21)$$

(iii) Large fish

When the fish is very large compared with the wave length ($L > 60\ \lambda$) , the back scattering of the fleshy part of the fish becomes important, since the fish body acts like a plane target. The total scattering cross section of the fish must be the sum of the scattering cross sections of the swimbladder, the backbone and the fleshy body. The contribution of the first two parts is given in equation (21).

The scattering cross section at the broadside aspect of a plate is determined by equation (22), where A is the area of the plate and β is the reflectivity.

$$\sigma_{broadside} = \frac{4\pi A^2}{\lambda^2} \beta^2 \qquad (22)$$

Consider now a fish swimming horizontally at a distance r from the sound source. At any position, the fleshy part of the fish can be replaced by a plate of area equal to geometrical cross section of the fish body and of reflectivity equal to that of fish flesh. This plate is normal to the plane determined by r and the axis from head to tail of the fish. The approximate sizes of the plate are:

length = $b = L_1 = 0.93\ L$

width = a : depending on the position of the fish, a varies between $H_1 = 0.195\ L$, and $B_1 = 0.112\ L$ (see parts of a fish). On average

$$a = \frac{H_1 + B_1}{2} = 0.153\ L\ .$$

Let ϕ be the incident angle, the scattering cross section of the fleshy body of the fish is:

$$\sigma_{F.B} = \sigma_{broadside} \times D(\phi) \qquad (23)$$

where $D(\phi)$ is the directivity factor:

$$D(\phi) = \frac{\sin k\ b\ (\sin \phi)}{k\ b\ (\sin \phi)} \cos^2 \phi \qquad (24)$$

The average value of this factor over various incident angles is

$$\overline{D}(\phi) = \frac{2}{\pi} \int_0^{\pi/2} \left[\frac{\sin\ (k\ b \sin\ )}{k\ b \sin}\right]^2 \cos^2\phi \ d\phi \qquad (25)$$

Let $u = k\ b \sin \phi$

$$D(\phi) = \frac{2}{\pi} \cdot \frac{1}{k\ b} \int_0^{k\ b} \left(\frac{\sin u}{u}\right)^2 \left(1 - \frac{u^2}{k^2 b^2}\right)^{\frac{1}{2}} du \qquad (26)$$

for large fish $kb = \frac{2\pi}{\lambda} \cdot b > 110\pi >> \pi$. The function $\left(\frac{\sin u}{u}\right)^2$ tends to zero very fast after $u > \pi$, while $\left(1 - \frac{u^2}{k^2 b^2}\right)^{\frac{1}{2}}$ decays slowly. Thus $\bar{D}(\phi)$ can be approximated by

$$D(\phi) = \frac{2}{\pi} \times \frac{1}{k\,b} \int_0^{\infty} \frac{\sin u}{u}^2 \, du = \frac{1}{k\,b} \qquad (27)$$

Hence the average scattering cross section of the fleshy body is

$$\bar{\sigma}_{F.B} = \frac{2a^2 b}{\lambda} \beta^2 \qquad (28)$$

Substituting the values of a and b in equation (28)

$$\overline{\sigma_{F.B}} = 8.43 \times 10^{-5} \frac{L^3}{\lambda} \quad \text{in fresh water} \qquad (29)$$

$$\overline{\sigma_{F.B}} = 1.57 \times 10^{-3} \frac{L^3}{\lambda} \quad \text{in salt water} \qquad (30)$$

The total back scattering cross section of the fish is then

$$\bar{\sigma} = 8.43 \times 10^{-5} \frac{L^3}{\lambda} + 6.14 \times 10^{-3} L^2 \quad \text{in fresh water} \qquad (31)$$

$$\bar{\sigma} = 1.57 \times 10^{-5} \frac{L^3}{\lambda} + 6.14 \times 10^{-3} L^2 \quad \text{in salt water} \qquad (32)$$

All the formulas given for the back scattering cross sections can be converted into the formulas for target strengths by using equation (2). Figures 5 and 6 show the back scattering cross section and target strength of fish as the functions of fish length, when the frequency is equal to 60 kHz. These curves are plotted by using the formulas (11), (12), (21), (31), (32) and (2). In comparison to the experimental results, say at dorsal aspect, the target strength of a fish, averaged over different incident angles, is from 5 to 10 dB lower.

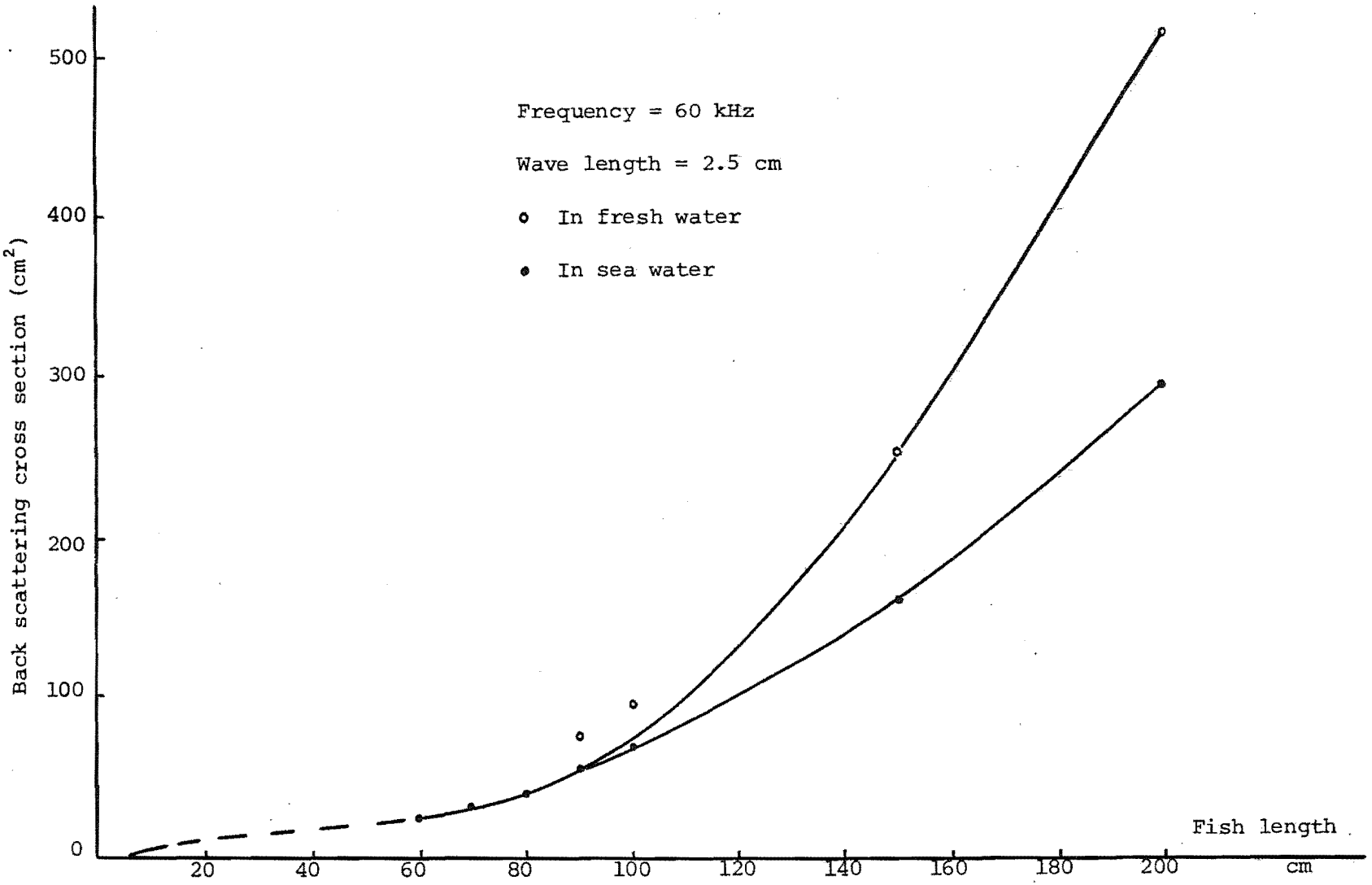


Fig. 5. Variation of the back scattering cross section with rspect to the length of fish.

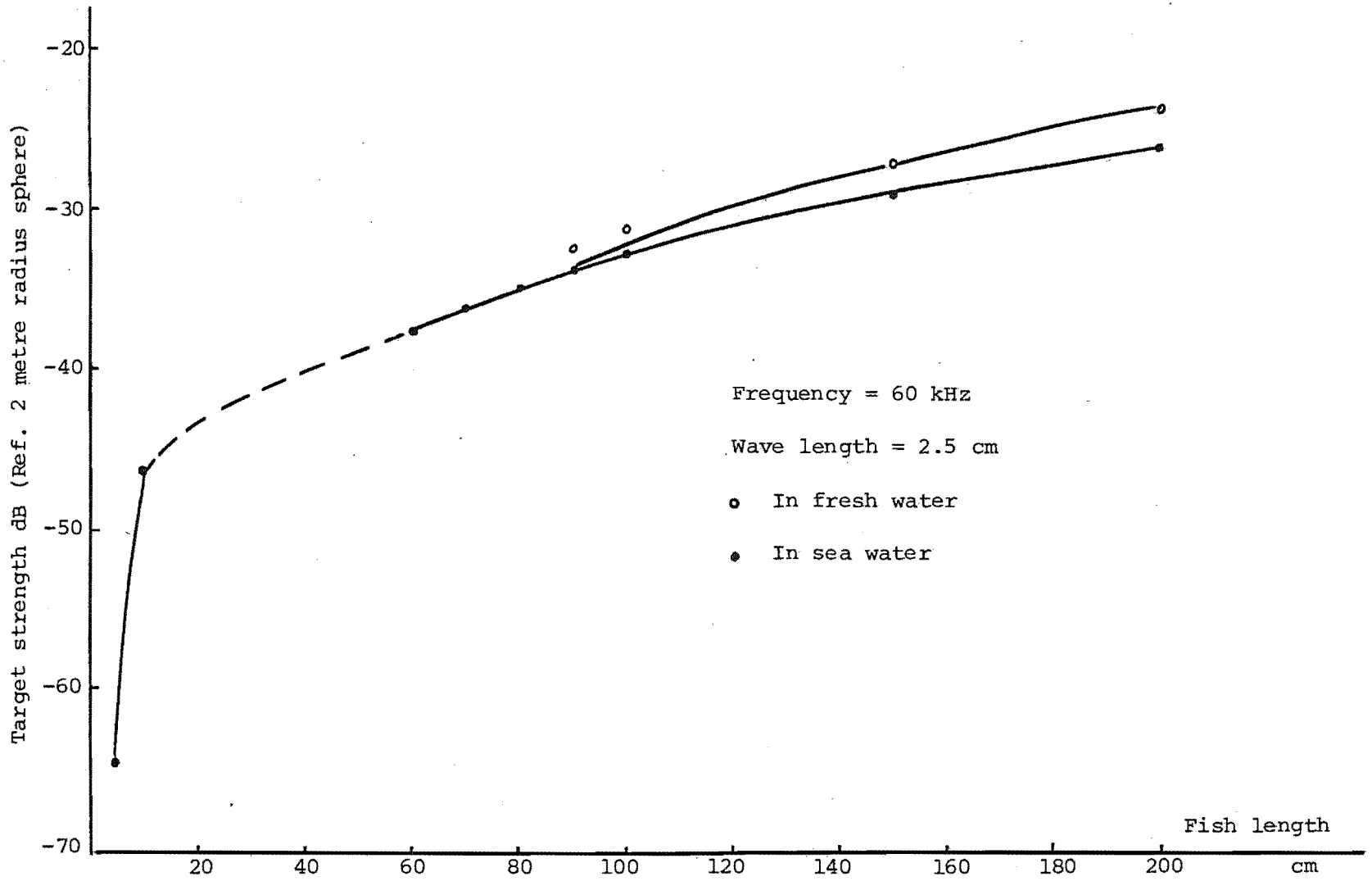


Fig. 6. Variation of the target strength with respect to the length of the fish.

7.4 REFLECTION FROM A SHOAL OF FISH

A shoal of fish is defined by the following characteristics:

(1) type of fish

(2) density, i.e. the number of fishes per unit volume

(3) population

(4) shape of the shoal.

The purpose of this section is to provide a method to estimate the scattering cross-section or target strength of a fish shoal. To simplify the problem, it will be assumed that the shoal is evenly distributed by fishes of equal size.

Considering the simplest case where the fish shoal is either of a flat shape, or sparsely distributed so that there is no overlapping of echoes from individual fish, thus the total back scattering cross section of the shoal is:

$$\sigma_t = n \bar{\sigma} \tag{33}$$

where $\bar{\sigma}$ is the average scattering cross-section of a fish, and n is the population or the total number of fish of the shoal. However, in actual existing volumetric shoals, fish are frequently so densely concentrated that the total echo from the shoal is more or less attenuated by absorption of ultrasonic energy in fish bodies. Shishkova [9] found that the coefficient of attenuation increases with the increasing concentration of fish, and presented this effect in Figure 8. To determine the scattering cross-section of a volumetric shoal the following method may be used.

Let $A(u)$ be the geometrical cross section of a shoal of fish at distance u from its front surface. Let N be the density of fish in the shoal, then the number of fish in the volume element $A(u) \cdot du$ is $N A(u)\, du$.

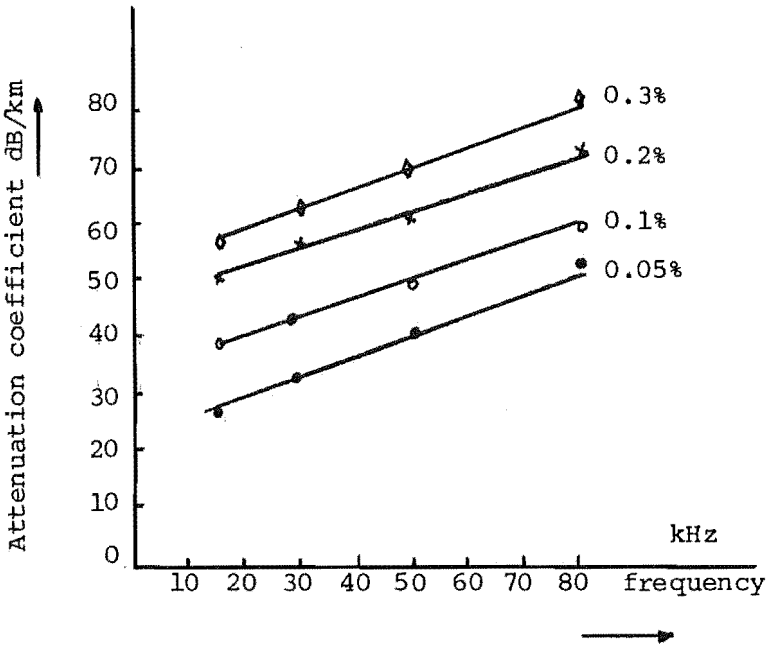


Fig. 7. Extinction of ultrasound in fish schools of different concentration in relation to frequency. The concentration is the biomass per cent of water volume.

(Reproduced from [9])

If there are no fish in front of this volume element (layer), its scattering cross section will be

$$\sigma_{LO}(u) = N\, A(u)\, \bar{\sigma}\, du \tag{34}$$

and its target strength is

$$T_{LO}(u) = 10 \log_{10} \frac{N\, A(u)\, \bar{\sigma}\, du}{4\pi} \tag{35}$$

Frequently the attenuation coefficient of sound in the fish shoal, α_f, is expressed in dB per unit length. Consequently, a two way transmission of sound from the front of the shoal to the layer of fish $A(u)\, du$ results in an attenuation of the echo signal by $2\alpha_f u$. The target strength of the layer $A(u)du$ becomes

$$T_L(u) = 10 \log_{10} \frac{N\, A(u)\, \bar{\sigma}\, du}{4\pi} - 2\, \alpha_f\, u \tag{36}$$

or

$$T_L(u) = 10 \log_{10} \frac{1}{4\pi} \left[N\, A(u)\, \bar{\sigma} \times 10^{-\frac{\alpha_f}{5} u}\, du \right] \tag{37}$$

The scattering cross section of this layer is

$$\sigma_L(u) = N\, A(u)\, \bar{\sigma} \times 10^{-\frac{\alpha_f}{5} u}\, du \tag{38}$$

Let d be the thickness of the shoal, then its total scattering cross section must be

$$\sigma_S = N\, \bar{\sigma} \int_0^d A(u)\, 10^{-\frac{\alpha_f}{5} u}\, du \tag{39}$$

If the shape of the shoal is known, $A(u)$ can be determined and the integration can be calculated. Suppose that $A(u)$ is constant, then

$$\sigma_S = N\, \bar{\sigma}\, A \int_0^d 10^{-\frac{\alpha_f}{5} u}\, du \tag{40}$$

$$\sigma_S = N\, \bar{\sigma}\, A \int_0^d \exp\left(-\frac{\alpha_f}{5} \ln 10\, u\right) du$$

$$\sigma_S = \frac{N \bar{\sigma} A}{0.46\ \alpha_f} \left(1 - e^{-0.46\ \alpha_f d}\right) \qquad (41)$$

Example 1. $N = 4$ fish/m^3 , $A = 25$ m^2, $d = 10$ m , $\bar{\sigma} = 50$ cm^2 (average scattering cross section of fish of about 90 cm length), $\alpha_f = 0.1$ dB/m . Using equation (41), the scattering cross section of the shoal is

$$\sigma_S = 40\,077 \text{ cm}^2$$

and its target strength, relative to a sphere of two 2 m radius, is

$$T.S = -5 \text{ dB}$$

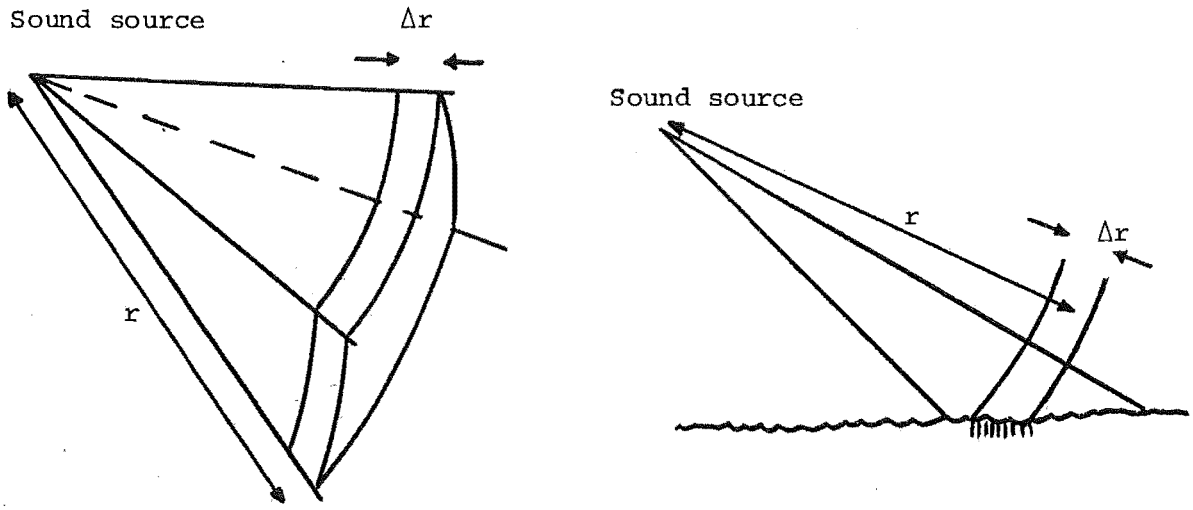
(If the shoal is flat, the total scattering cross section of the shoal is $50\,000$ cm^2 , and $T.S \simeq -4$ dB.)

7.5 CHARACTERISTICS OF NOISE DUE TO SEA REVERBERATION IN CTFM SONAR

It was first shown by Kay [10] and Smith [11] that the audible frequency noise due to sea reverberation produced by a CTFM sonar, within a critical bandwidth [12], is statistically identical to band limited white noise. This suggestion is applicable only under some conditions and to some extents as follows.

Two types of reverberation which are frequently encountered in forward looking sonars: (i) volume reverberation, and (ii) bottom reverberation, are considered. Assuming that the beam pattern is idealized and considering an annular region, Δr , at the range r , containing n scatterers (Figure 8), in both cases of volume and bottom reverberation the received signal can be represented by

$$V(t) = \sum_{i=1}^{n} a_i S(t - \tau_i) \qquad (42)$$



a) Volume reverberation b) Bottom reverberation

Fig. 8. Typical types of reverberation in forward looking sonar.

where $s(t)$ is the transmitted signal, a_i and τ_i are the random variables respectively related to the strength and the time delay of the returned signal from the i-th scatterer.

$$S(t) = \exp\left[j2\pi(f_2 t_s - m t_s^2) \right] \qquad (43)$$

$$0 \leq t_s \leq T_s$$

The parameters, f_2 , m , T_s , etc. were defined in Chapters 1 and 2. The audible noise produced by the demodulation of $V(t)$ and $S(t)$ is given by

$$N(t) = S(t)\, V(t) = \sum_{i=1}^{n} a_i \exp[j2\pi(m\tau_i t_s - f_2 \tau_i)] \qquad (44)$$

Therefore, if the following assumptions are valid:

(i) The scatterers are evenly distributed with the annular region Δr ,

(ii) The strengths of the scatterers are statistically equal.

then in stationary conditions, $N(t)$ is a band limited pseudo-random

noise signal having equal spacing and equal amplitude components, $f_{ai} = m \tau_i$, $i = 1$, n .

In dynamic conditions,tthe scatterers at two sides of the sonic beam have the relative range rates lower than that near the median of the beam. The frequency components of the noise signal due to reverberation,

$$f_{ai} \simeq m \tau_i - (f_2 - m t_s) \tau_i \qquad (45)$$

become unevenly spaced. However, if the number of scatterers in the annular region Δr is large enough so that the spacing between two successive frequency components is smaller than, say, 2 Hz (the minimum perceptible change in frequency [10]) the characteristics of noise due to sea reverberation under dynamic conditions will be similar to that under stationary conditions. As the boat moves about, the scatterers in a certain annular region Δr move towards the sonar system, and the scatterers in the annular vicinity $\Delta r'$ will move in to fill up the region Δr at the same time. Therefore the total background noise produced by the CTFM sonar is heard as to be stationary, providing that no large individual scatterers are present in the sonic beam.

Using the fishing sonar designed by Smith [11] (the system has four different range codes of 53.3 Hz/m, 26.7 Hz/m, 13.3 Hz/m and 6.7 Hz/m for four different maximum ranges of 94, 188, 375 and 750 m), the research group led by Kay (University of Canterbury, New Zealand) during the period since September 1975 have carried out many tests on target detection at Lake Coleridge (200 m of water) and measured the noise due to sea reverberation at Tasman Bay (shallow, 20 m of water). Figure 11 compares the noise due to sea reverberation in dynamic conditions and the pseudo random noise, while Figures 9 and 10 show the increase in noise level when the sonic beam hits the sea bed.

Fig. 9.-Noise due to sea reverberation recorded at Tasman bay, N.Z. (20m of water, sandy bottom). Transducers are of vertical beam width of 25 deg., and tilt 10 deg. Range code = 53.3 Hz/m. Result is analysed on the frequency scale of 10 Hz/cm, over different frequency bands centered at: (a) 800 Hz, (b) 2000 Hz, (c) 2400 Hz, (d) 2800 Hz, (e) 3300 Hz, and (f) 4300 Hz.

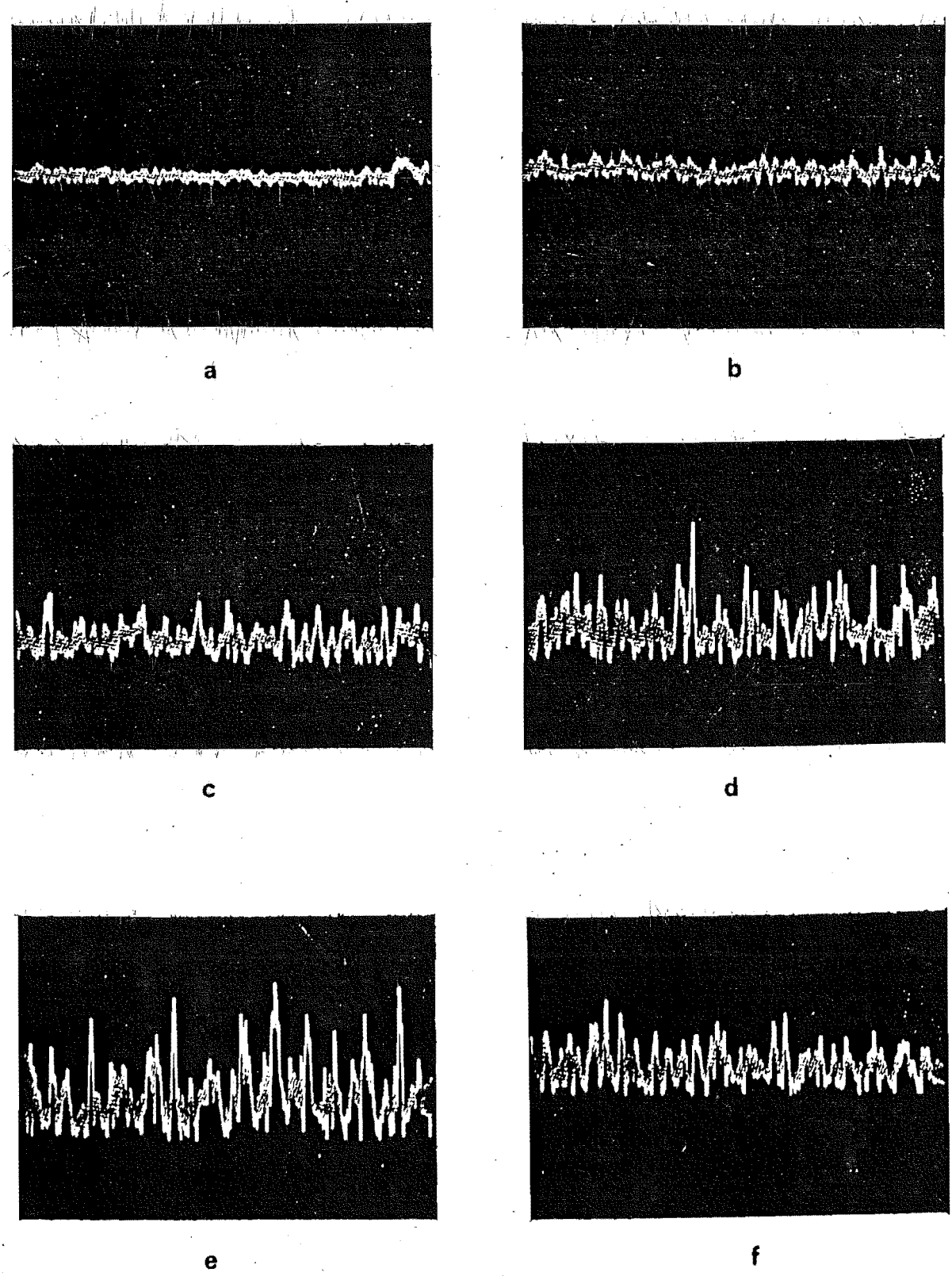


Fig. 10.-Spectra of noise due to sea reverberation in C.T.F.M. sonar. Range code is 53.3 Hz/m. Reference frequency = 0 Hz, scale = 1 kHz/cm.
(a) Top: tilt angle = 5 deg. bottom: tilt angle = 10 deg.
(b) Top: tilt angle = 15 deg. bottom: tilt angle = 20 deg.

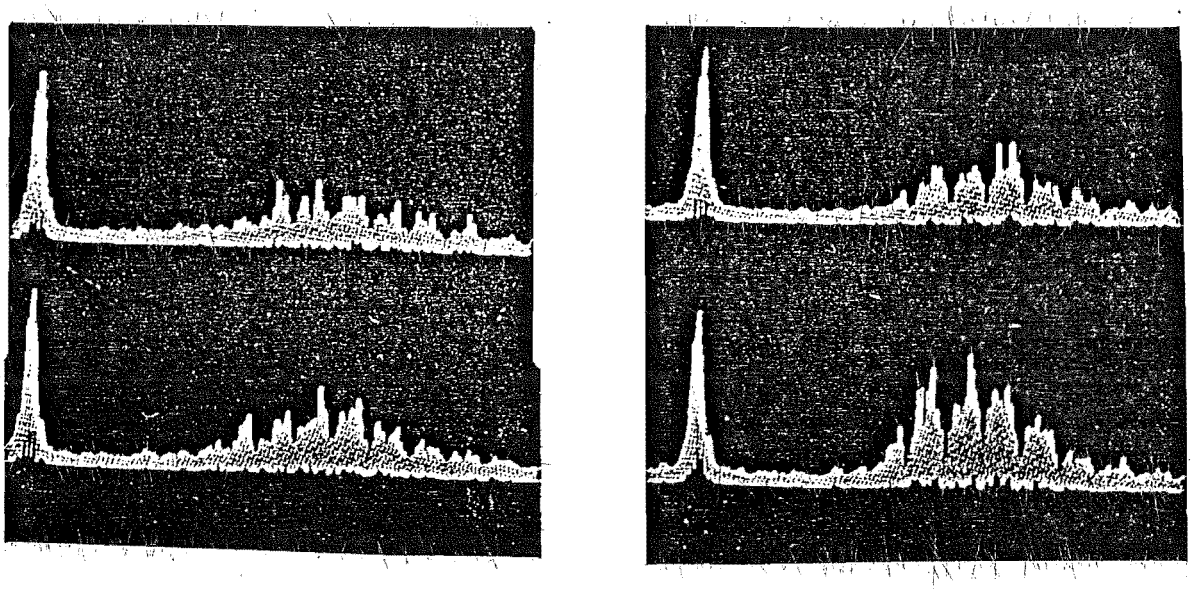
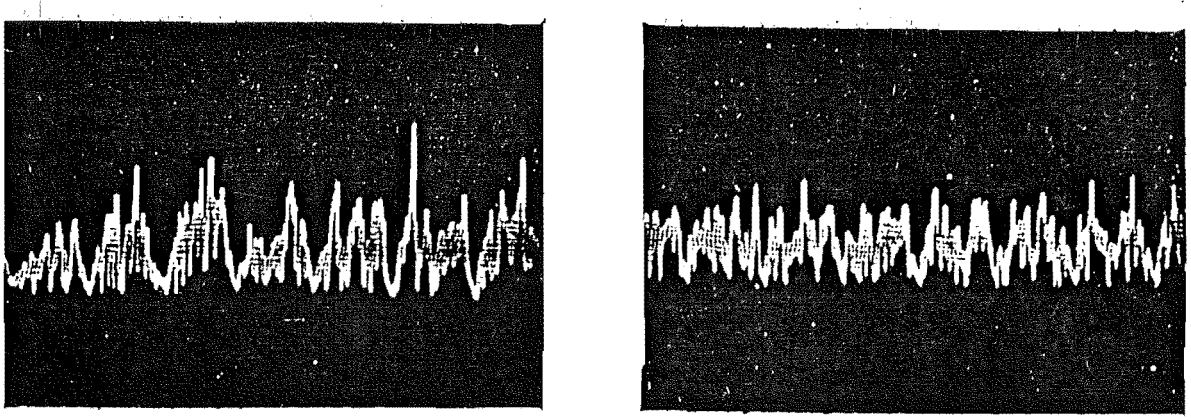


Fig. 11.-Comparison of noise due to sea reverberation and pseudo-random noise.
(a) Noise due to sea reverberation in dynamic condition. Range code = 26.7 Hz/m, center frequency = 3000 Hz, scale = 10 Hz/cm.
(b) Pseudo-random noise, spacing between two successive components is 1 Hz, center frequency = 3000 Hz, scale = 10 Hz/cm.



a b

7.6 REVERBERATION LEVEL AND POWER SPECTRAL DENSITY OF AUDIBLE NOISE IN CTFM SONAR

It is seen in section 7.5 that the spectrum of audible noise due to sea reverberation is typically of the form (Fig. 12)

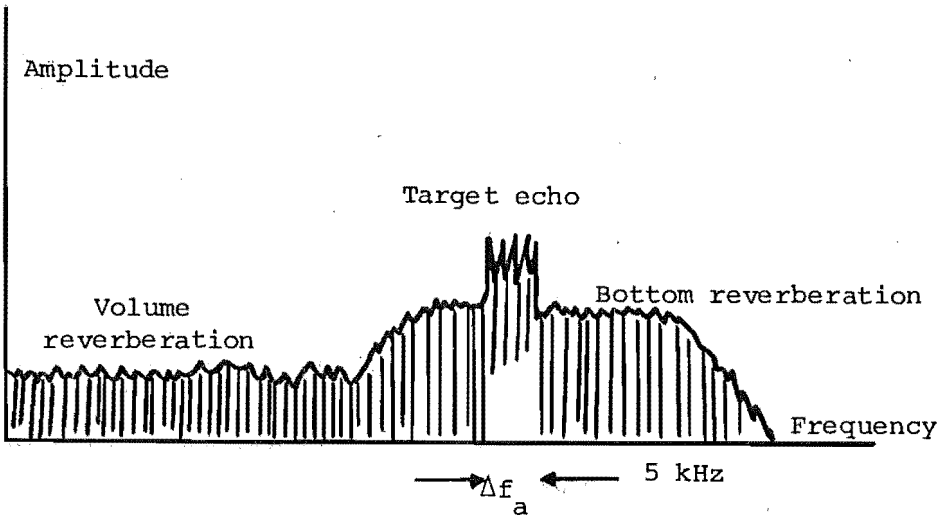


Fig. 12. A typical noise spectrum due to volume and bottom reverberation. Depending on the depth of water, more or less, bottom reverberation is presented. Echo signal is of a bandwidth Δf_a .

Since a finite target extending over an arbitrary range increment Δr is displayed by an audible signal of a bandwidth equal to

$$\Delta f_a = \frac{2m}{C} \Delta r \qquad (46)$$

where $\frac{2m}{C}$ is the range code of the CTFM sonar. The least portion of noise which effectively masks the signal is then determined by the reverberation from the scatterers contained in the annular region Δr . The purpose of this section is to provide a method to determine the sound intensity at the receiver due to reverberation from either volume scatterers or bottom scatterers occupying any arbitrary annular region Δr . In other words, to determine the power spectral density of noise due to sea reverberation.

Let the beam patterns of the projector and the hydrophone be denoted by $P(\theta,\psi)$ and $H(\theta,\psi)$ respectively, and the axial intensity of the transmitted wave at unit distance be I_o. The source level is defined by

$$S.L = 10 \log_{10} I_o \tag{47}$$

7.6.1 Volume Reverberation

Considering an elemental volume, dV, of volume scatterers determined by the elemental solid angle $d\Omega$, and Δr, at a distance r (Fig. 13)

$$dV = r^2 \, d\Omega \, \Delta r \tag{48}$$

Let S_v be the ratio of the intensity of the back scattering produced by a unit volume, at a unit distance from the volume dV, to the intensity of the incident sound wave. The sound intensity reverberated from the volume dV, at the hydrophone is

$$\frac{I_o \, S_v}{r^4} \, P(\theta,\psi) \, H(\theta,\psi) \, dV \tag{49}$$

Substituting equation (48) in (49), the total reverberation level due to back scattering from the annular region Δr is given by

$$R.L_v = 10 \log \left(\frac{I_o \, S_v}{r^2} \Delta r \int_{\theta,\psi} P(\theta,\psi) \, H(\theta,\psi) \, d\Omega \right) \tag{50}$$

where

$$\Psi = \int_{(\theta,\psi)} P(\theta,\psi) \, H(\theta,\psi) \, d\Omega \tag{51}$$

is constant and interpreted as the total equivalent beam width of the projector-hydrophone combination (see Appendix 6). Thus

$$R.L_v = 10 \log \left(I_o \cdot \Psi \cdot S_v \, \frac{\Delta r}{r^2} \right) \tag{52}$$

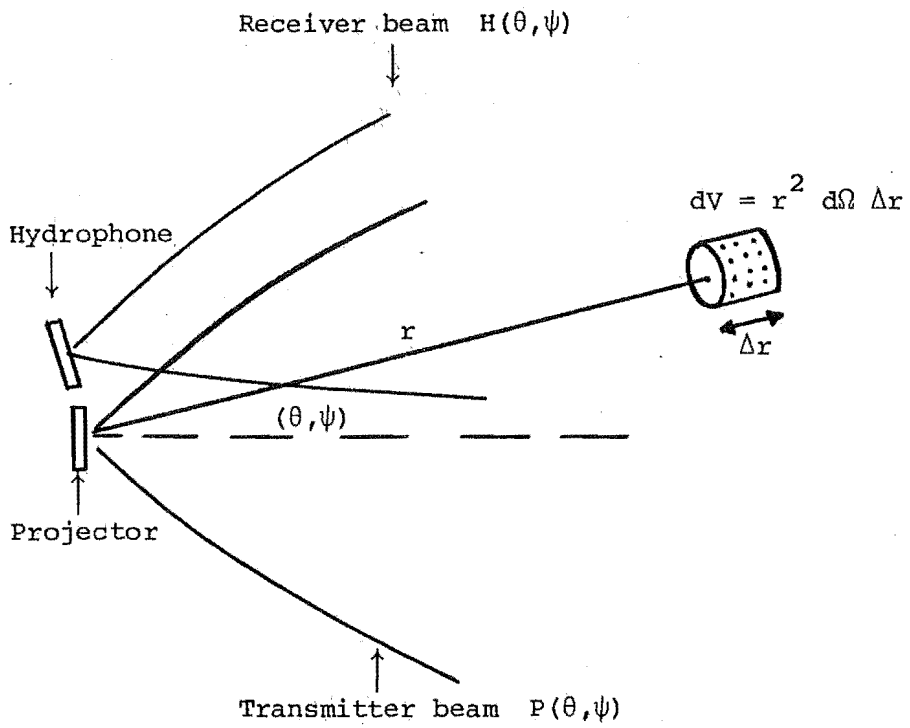


Fig. 13. Geometric view of volume scattering.

Let K be a constant proportional to the mean squared voltage response of the hydrophone, and f_a be the range coding frequency. The noise level due to this amount of reverberation is

$$N.L_v = 10 \log \left(K I_o \cdot \Psi \cdot S_v \cdot \frac{2m}{C} \quad \frac{\Delta f_a}{f_a^2} \right) \qquad (53)$$

The power spectral density of noise due to volume reverberation is then

$$N.D_v = 10 \log \left(K I_o \cdot \Psi \cdot S_v \quad \frac{2\,m}{C\, f_a^2} \right) \qquad (54)$$

which decreases with respect to frequency at a rate of 6 dB/oct , and is entirely deterministic whenever the volume scattering strength, $10 \log S_v$, is known. The typical scatterers producing volume reverberation are the marine life and inanimate matter distributed in the sea and the inhomogenous structure of the sea itself. Barakos [13] showed that the volume scattering strength (measured at 24 kHz) decreases with depth at a rate of 5 dB/1000 ft (equivalent to 1.67 dB/100 m) from the value of -77 dB (near the surface). An increase of 5 to 15 dB of the scattering strength is observed when there exists a deep sea layer (Figure 14). These layers were believed to be of biological origin [14,15,16], remaining at the depths between 300 and 900 m during the daytime, and migrating to the surface at night. The scattering from these layers is only significant at the frequencies lower than 40 kHz (resonant frequencies of small fish) [17]. The variation of the volume scattering strength with respect to frequency (when no deep sea layers exist) is shown in Figure 15 [18]. Within the frequency band from 40 to 80 kHz, the variation of S_v is very minor, and $10 \log S_v \simeq -92$ dB .

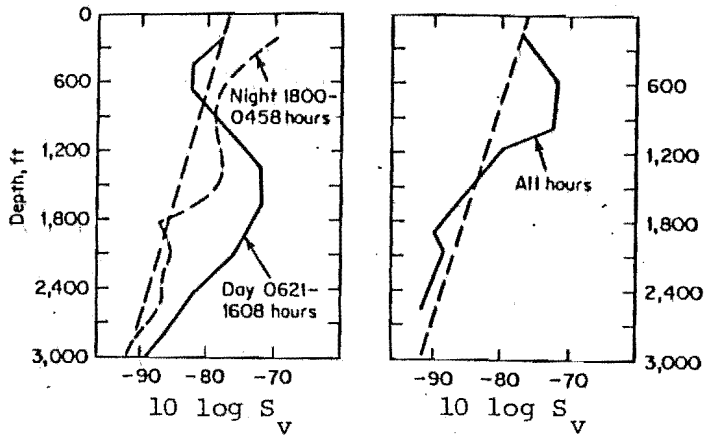


Fig. 14. Volume scattering strength at 24 kHz versus depth as measured in two Pacific Ocean areas: (a) Guadalupe Island area, latitude 29°N , (b) Queen Charlotte Island area, latitude 51°N. Dashed lines are estimated minimum values [19].

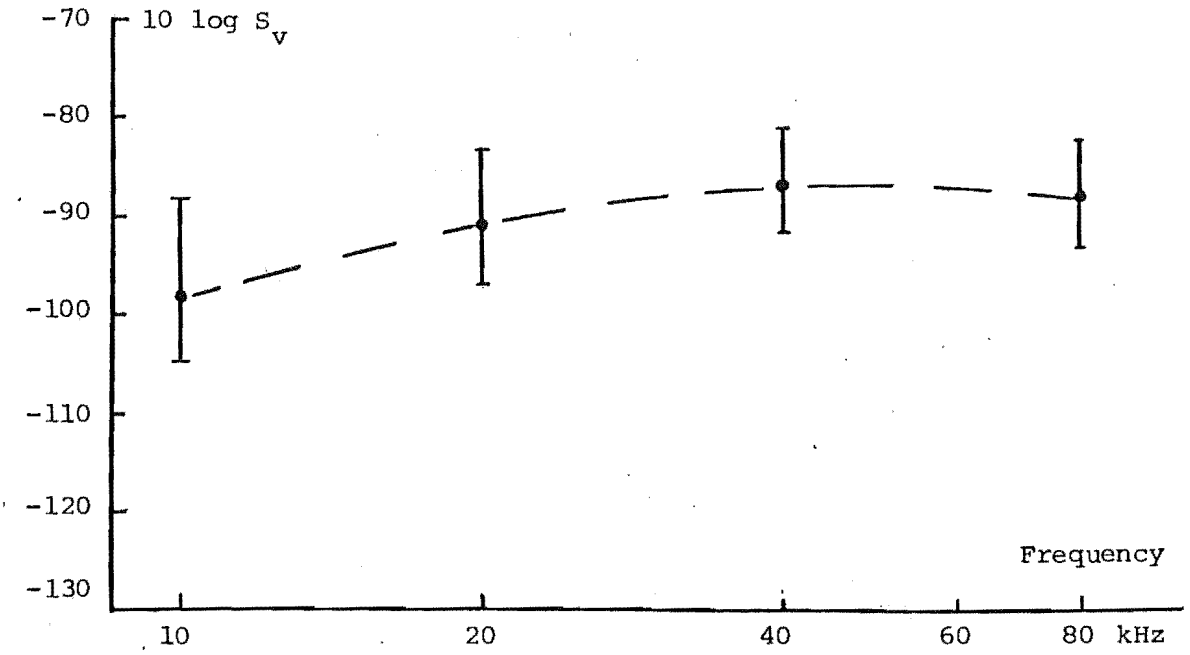


Fig. 15. Variation of volume scattering strength with frequency, averaged from data at nine locations at depth less than 250 ft. [18].

7.6.2 Bottom Reverberation

Considering an elemental area, dA, of the scattering bottom at the position (r, θ, ψ) as described in Figure 16, where 0 is the position of the sonar system, r is the distance from 0 to dA, θ and ψ are respectively the azmiuth and elevation of dA.

$$dA = r \Delta r \, d\theta \tag{55}$$

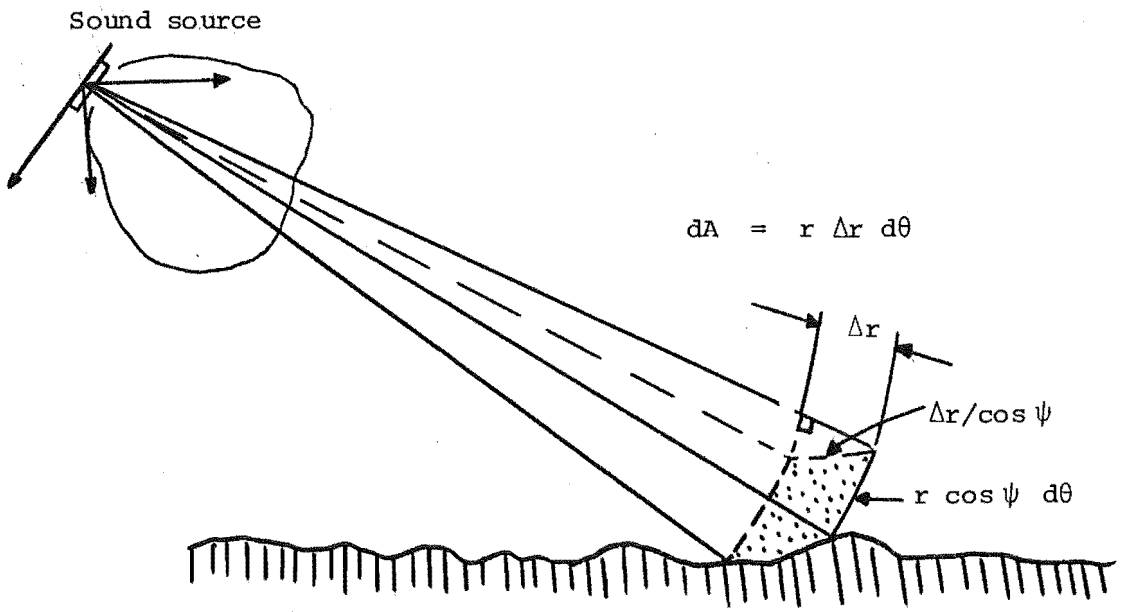


Fig. 16. Geometric view of bottom reverberation.

Let $S_B(\psi)$ be the ratio of the intensity of the back scattering produced by a unit area, at a unit distance from the area dA, to the intensity of the incident wave. $S_B(\psi)$ is a function of the incident angle ψ and depends on the roughness as well as the structure of the sea bed. Using similar manipulation, the total reverberation level contributed by the bottom scatterers in the annular region Δr is given by

$$R.L_B = 10 \log \left(\frac{I_o}{r^3} \Delta r \, S_B(\psi) \int_\theta P(\theta,\psi) \, H(\theta,\psi) \, d\theta \right) \qquad (56)$$

Now, if $P_1(\theta)$, $P_2(\psi)$ and $H_1(\theta)$, $H_2(\psi)$ are respectively horizontal and vertical beam-patterns of the projector and the hydrophone, $P(\theta,\psi)$ and $H(\theta,\psi)$ can be approximated by:

$$P(\theta,\psi) \simeq P_1(\theta) \, P_2(\psi) \qquad (57)$$

$$H(\theta,\psi) \simeq H_1(\theta) \, H_2(\psi) \qquad (58)$$

Then

$$\int_\theta P(\theta,\psi) \, H(\theta,\psi) \, d\theta = P_2(\psi) \cdot H_2(\psi) \cdot \Theta \qquad (59)$$

where

$$\Theta = \int^{2\pi} P_1(\theta) \, H_1(\theta) \, d\theta \qquad (60)$$

is a constant representing the horizontal combination beam width (see Appendix 6). Let ϕ be the width of the vertical beam patterns, and suppose that they are idealized

$$P_2(\psi) = H(\psi) = 1 \quad \gamma - \frac{\phi}{2} \leq \psi \leq \gamma \quad \frac{\phi}{2}$$

$$= 0 \quad \text{elsewhere} \qquad (61)$$

where γ is frequently called the tilt angle of the transducer.

Substituting equations (59), (60) and (61) in equation (55), we have

$$R.L_B = 10 \log_{10} \left(I_o \cdot \Theta \cdot S_B(\psi) \cdot \frac{\Delta r}{r^3} \right) \qquad (62)$$

$$\gamma - \frac{\phi}{2} \leq \psi \leq \gamma + \frac{\psi}{2}$$

According to equation (62), the bottom reverberation is a function of r and ψ . However, if the vertical beam width ϕ is not too large, say, less than 30^o, then $S_B(\psi)$ can be considered as constant. Many workers [19 - 24] measured the bottom scattering strength,

$10 \log S_B$, over a large range of frequency, at different grazing angles, and for different types of sediment; and came to the similar conclusion that:

(i) There appears to be no significant frequency dependence of bottom back scattering over a wide range, say, 7 octaves.

(ii) There is a wide variation of the bottom scattering strength over different types of sea bed. Frequently, at grazing angles smaller than, say, 60°, the hard bottoms have a much higher scattering strength than the soft bottoms.

(iii) The variation of the scattering strength with respect to grazing angle can be described by Lambert's law for diffuse scattering from a rough surface (one constant) [28],

$$10 \log S_B = - u + 10 \log \sin^2 \psi \qquad (63)$$

or may be better fitted with Burstein and Keane's equation [23] (Two constant).

$$10 \log S_B = - a + b \log \tan \psi \qquad (64)$$

where the constant u , a and b are related to the types of the sea bed. The data given in Figure 17 show that the variation of the scattering strength, as $\psi < 60^\circ$, is very little in comparison to its variation over different types of sea bed. The conditions of $\psi < 60^\circ$ are frequently met in forward looking sonars. Consequently, if the vertical beam width ϕ is not too large, say, more than 30°, the approximation of S_B to a constant in equation (62) is reasonable. If D is the distance from the transducers to the sea bottom, the variable ψ can be replaced by

$$\psi = \sin^{-1} \frac{D}{r} \qquad (65)$$

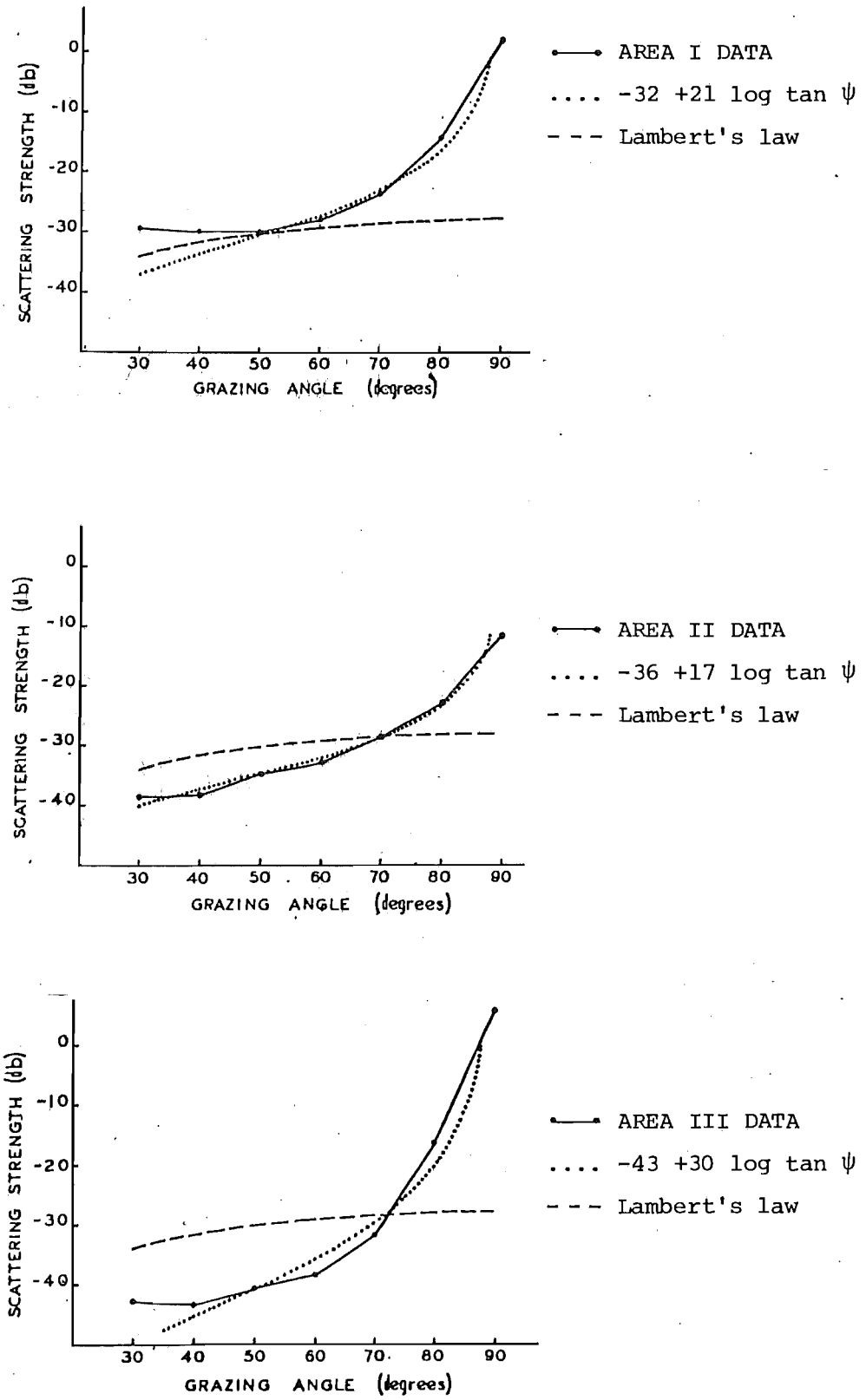


Fig. 17. Bottom scattering strength plotted as a function of grazing angle ψ. The bottom at area I (1500 fathoms of water) and at area III (2050 fathoms of water) is composed of mud. The bottom at area II (350 fathoms of water) appears to be coral [23].

Hence the reverberation level contributed by the annular region r is

$$R.L_B = 10 \log \left(I_o \cdot \Theta \cdot S_B \cdot \frac{\Delta r}{r^3} \right) \qquad (66)$$

where $\frac{D}{\sin(\gamma + \frac{\phi}{2})} \leq r \leq \frac{D}{\sin(\gamma - \frac{\phi}{2})}$

The corresponding noise level is

$$N.L_B = 10 \log \left(K \cdot I_o \cdot \Theta \cdot S_B \cdot \left(\frac{2m}{C}\right)^2 \frac{\Delta f_a}{f_a^3} \right) \qquad (67)$$

where $\frac{2m}{C} \frac{D}{\sin(\gamma + \frac{\phi}{2})} \leq f_a \leq \frac{2m}{C} \frac{D}{\sin(\gamma - \frac{\phi}{2})}$

And the noise density is

$$N.D_B = 10 \log \left\{ \frac{K \cdot I_o \cdot \Theta \cdot S_B}{f_a^3} \left(\frac{2m}{C}\right)^2 \right\} \qquad (68)$$

This decreases with respect to frequency at a rate of 9 dB/oct. Data given by McKinney and Anderson [24] are reproduced in Figure 18 for use in calculating the reverberation or noise level.

7.7 SONAR EQUATION AND MAXIMUM DETECTABLE RANGE

Considering a finite target situated at a distance r from the source and extended over an interval Δr. Its target strength, as defined in section 7.2, is

$$T.S = 10 \log \frac{I_{r1}}{I_i}$$

If θ_o and ψ_o are the azimuth and elevation angles of the target, the level of the signal returned from the target to the hydrophone will be

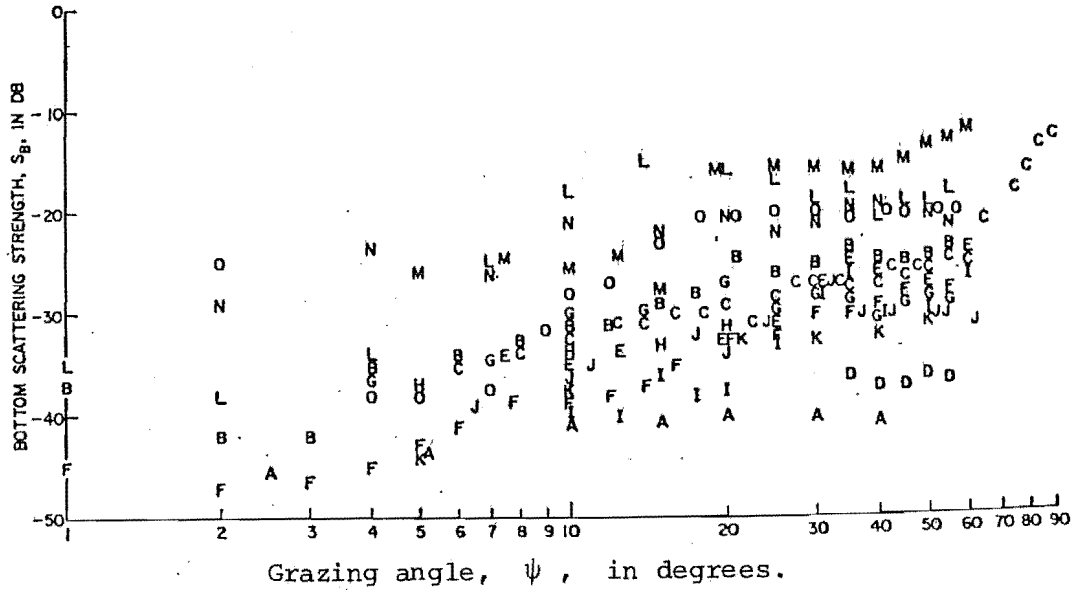


Fig. 18. Bottom scattering strength as a function of grazing angle, ψ, for a frequency of 100 kHz. Symbols are the same as given in Table II. (Reproduced from [24])

Table II. Description of different types of sediment.

Area	Sediment description	Mean diam d	d/λ	Percent Sand	Silt	Clay	Gravel and shell
A	Fine sandy mud	0.002 mm	0.00013	26	74		...
B	Medium sandy clay	0.011	0.00074	35	16	48	...
C	Muddy fine sand	0.013	0.00087	40	28	32	...
D	Very fine sandy mud	0.017	0.00113	46	31	23	...
E	Silty fine sand	0.068	0.00454	69	19	8	4
F	Clayey medium sand	0.10	0.00667	72	10	17	...
G	Muddy fine sand	0.148	0.00987	84	8	8	...
H	Mud-Sand	0.23	0.0154	90	5	...	5
I	Sand	0.24	0.016	97	2	1	...
J	Medium sand	0.4	0.0266	91	1	...	8
K	Medium sand	0.4	0.0266	...	...	...	...
L	Sandy pebble gravel	3.2	0.213	33	2		65
M	Sandy pebble gravel	4.2	0.28	28	1	1	70
N	Solid rock	...	...	...	...	...	...
O	Solid coral, huge chunks coral growth	...	...	94	2	2	...

$$E.L = 10 \log \left(\frac{I_o}{r^4} P(\theta_o,\psi_o)\ H(\theta_o,\psi_o)\ \frac{I_{rl}}{I_i} \right) \qquad (69)$$

or $$E.L = 10 \log \left(\frac{I_o}{r^4} P(\theta_o,\psi_o)\ H(\theta_o,\psi_o) \right) + T.S \qquad (70)$$

E.L is called the echo-level.

Since the target is extended or is an interval Δr , at least the echo signal is interfered with by the reverberation from all the scatterers inside the annular region Δr occupied by the target. Depending on whether the sonic beam hits the sea bottom at the range r , or not, either equation (52) or equation (66) will be used to calculate the reverberation level. Supposing that the target is near the axis of the transducer beam so that

$$P(\theta_o,\psi_o) \simeq H(\theta_o,\psi_o) \simeq 1 \qquad (71)$$

Then the maximum echo to reverberation ratio (in dB scale, it is equal to the difference EL - RL) is given by either

$$E.L - R.L = T.S - 10 \log \Psi - 10 \log (S_v \cdot r^2 \cdot \Delta r) \qquad (72)$$

(volume reverberation)

or

$$E.L - R.L_B = T.S - 10 \log \Theta - 10 \log (S_B \cdot r \cdot \Delta r) \qquad (73)$$

(bottom reverberation)

At the output of the demodulator, the signal level given by the target is

$$S.L = 10 \log \left(K \left(\frac{2m}{C}\right)^4 \frac{I_o}{f_a^4} \right) + T.S \qquad (74)$$

Therefore the signal to noise ratio over the bandwidth of the signal is either

$$S.L - N.L_v = T.S. - 10 \log \Psi - 10 \log \left(S_v \cdot f_a^2 \cdot \Delta f_a \cdot \left(\frac{C}{2m}\right)^3\right) \qquad (75)$$

(volume reverberation)

or

$$S.L - N.L_B = T.S. - 10 \log \Theta - 10 \log(S_B \cdot f_a \cdot \Delta f_a \cdot (\frac{c}{2m})^2) \quad (76)$$

(bottom reverberation)

If the target is so small that its coding sound is a pure tone, equations (75) and (76) become

$$S.L - N.D_V = T.S - 10 \log \Psi - 10 \log (S_V \cdot f_a^2 \cdot (\frac{c}{2m})^3) \quad (77)$$

and

$$S.L - N.D_B = T.S - 10 \log \Theta - 10 \log (S_B \cdot f_a \cdot (\frac{c}{2m})^2) \quad (78)$$

where $N.D_V$ and $N.D_B$ were defined previously as the noise densities due to volume reverberation and bottom reverberation. (S.L - N.D) is expressed in dB per Hz.

Thus, provided that the output signal of the CTFM demodulator is analysed on the band pass filters whose widths are as narrow as the displayed signal band width, the criterion for a target to be detectable is that its signal to noise ratio (per signal band width) is, at least, greater than 0 dB.

Example 2. A fish shoal having target strength $T.S = -7$ dB (about 500 fish, each of individual strength -34 dB) extends over $\Delta r = 5$ m.

(i) Deep water. The volume scattering strength is $10 \log_S S_V = -80$ dB . The total equivalent beam width of the projector-hydrophone combination is $\Psi = 0.1832$ rad^2 , or $10 \log \Psi = -7.37$ dB (see Appendix 6). Solving equation (72), the maximum range that the shoal is just detectable is $r_{max} = 4667$ m. Even in the worst case, $10 \log S_V = -70$ dB, $r_{max} = 1476$ m .

(ii) Shallow water. The horizontal combination beam width is $\Theta \simeq 24^o$ or 0.423 rad , $10 \log \Theta = -3.74$ dB (see Appendix 6).

If the beam hits a sandy bottom having scattering strength of $10 \log S_B = -30$ dB at a range shorter than the target range then the maximum detectable range of the shoal given by equation (73) is $r_{max} = 94$ m . However, if the system is operated at an area of muddy bottom, $10 \log S_B \simeq -40$ dB , the same shoal of fish can be detected at a maximum range of 940 m .

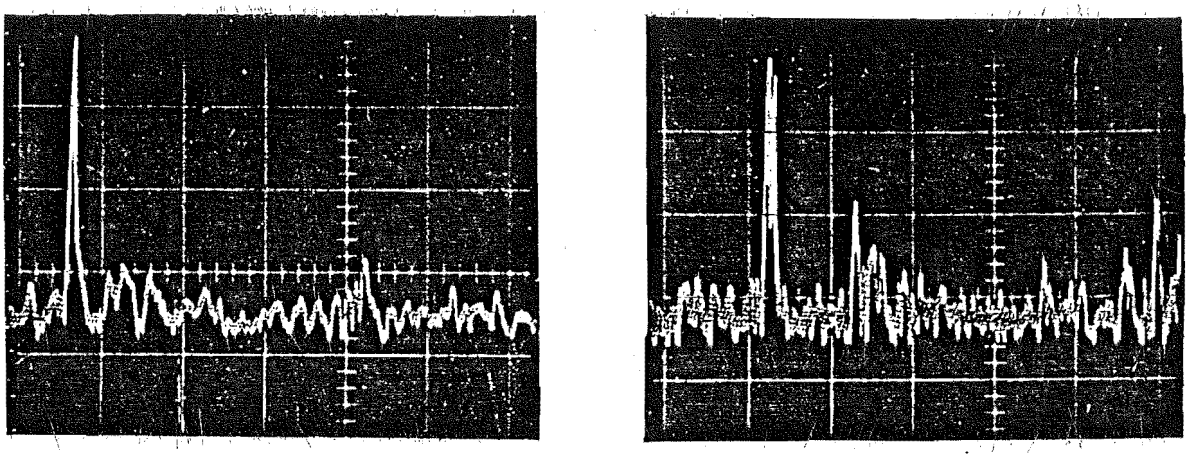
SOME EXPERIMENTAL RESULTS

Since 1975, a small group of researchers from the University of Canterbury (Kay et al.) has carried out many trials at different areas of New Zealand water, using the binaural fishing sonar described in chapter 1. The audible sound displayed by this sonar was recorded on magnetic tapes. A few samples of the recorded sound were analysed. These results are as follows:

1) In a trial at Lake Coleridge (200 metres of water) it was found that, at 40 m depth, a 20 in. steel sphere could still be detected at a range of almost 600 metres. A sample of the recorded sound was analysed and its result is shown in Figure 19. The 3345 Hz tone given by reflection from the sphere could be heard clearly and is comparable to a single tone of the same frequency. The sonar was operated at the range code of 6.67 Hz/m in this experiment. Thus the sphere was detected at a range of 517 metres.

The target strength of the sphere is $T.S = -18$ dB. If the volume scattering strength is $10 \log S_v = -85$ dB and the total equivalent beam width of the projector-hydroplane combination in dB is $10 \log \Psi = -7.37$ dB (see Appendix 6), using equation (77) the (S/N) ratio per Hz given by the sphere will be

$$S.L - N.D_v = 28.6 \text{ dB per Hz}$$



a b

Fig. 19. Experiment at Lake Coleridge, N.Z. (200 m of water), 20 in. diameter steel sphere detected by binaural sonar at range of 517 m. Range code = 6.7 Hz/m. Analysis shows tone detected at 3345 Hz.
(a) Center frequency = 4000 Hz, scale = 200 Hz/cm
(b) Center frequency = 3400 Hz, scale = 20 Hz/cm

The critical bandwidth of the ear at the frequency of 3400 Hz is about 550 Hz [25]. Hence the least (S/N) ratio per Hz required, so that the 3400 Hz tone is just audible, is 27.4 dB.

2) At Tasman Bay (22 m of water, sandy bottom), it was found that a triplane of 18 in. edge and -5.9 dB target strength could not be detected at a range further than 70 m.

The sonar was operated at the range code of 26.67 Hz/m. At $r = 65$ m, $f_a = 1734$ Hz. If the bottom scattering strength is $10 \log S_B = -30$ dB, and the horizontal combination beam width Θ in dB is $10 \log \Theta = -3.74$ dB (Appendix 6), using equation (78), the (S/N) ratio per Hz given by the tri-plane is

$$S.L - N.D_b = 24 \text{ dB per Hz}$$

According to the critical band theory, 24 dB is the least (S/N) ratio per Hz required so that the 1734 Hz tone is just audible. It should be noted that the vertical beam width is equal to 25°.

If the sonar is tilted at an angle of 10^{o}, the sonic beam will hit the bottom at $r \geq 58$ m.

3) An artificial shoal of fish was designed and tested at Lake Coleridge. The simulator is formed by forty hollow spheres of 11.5 cm diameter, which are hung on a 10 m length frame made of aluminium tubing (Figure 20). The cross section of each sphere is 104 cm^2 and is equivalent to that of five average fish (60 cm long). So the simulator is equivalent to a shoal of about 200 fish. The total target strength is T.S = - 14.8 dB. To sink this artificial shoal into water, each sphere was filled with an amount of lead shot which is just heavy enough to sink the sphere. The lead shot occupies only one-thirteenth of the hollow of the sphere.

The artificial shoal was sunk into water at a depth of 24 m, and the sonar system was operated at the range code of 26.67 Hz/m. The maximum detectable range of the shoal was about 150 m (a faint tone of average frequency about 4000 Hz could be heard and is comparable to a pure tone of the same frequency. Note that the maximum operating range of the sonar was only 180 m at the range code of 26.67 Hz/m.) Figures 21(a) and (b) show the spectra of the audible signal given by the shoal target at the ranges of 120 m (audible frequency $\simeq$ 3200 Hz) and 90 m (audible frequency $\simeq$ 2400 Hz) respectively. The bandwidth of the signal is from 260 Hz to 300 Hz. Applying equation (75), the (S/N) ratio over the signal bandwidth is found to be:

$$S.L - N.L_{v} = 26 \text{ dB} \quad \text{at 120 m range}$$

$$= 24 \text{ dB} \quad \text{at 150 m range.}$$

These data suggest that the artificial shoal has a target strength large enough to be detectable at the maximum operating range of the sonar system.

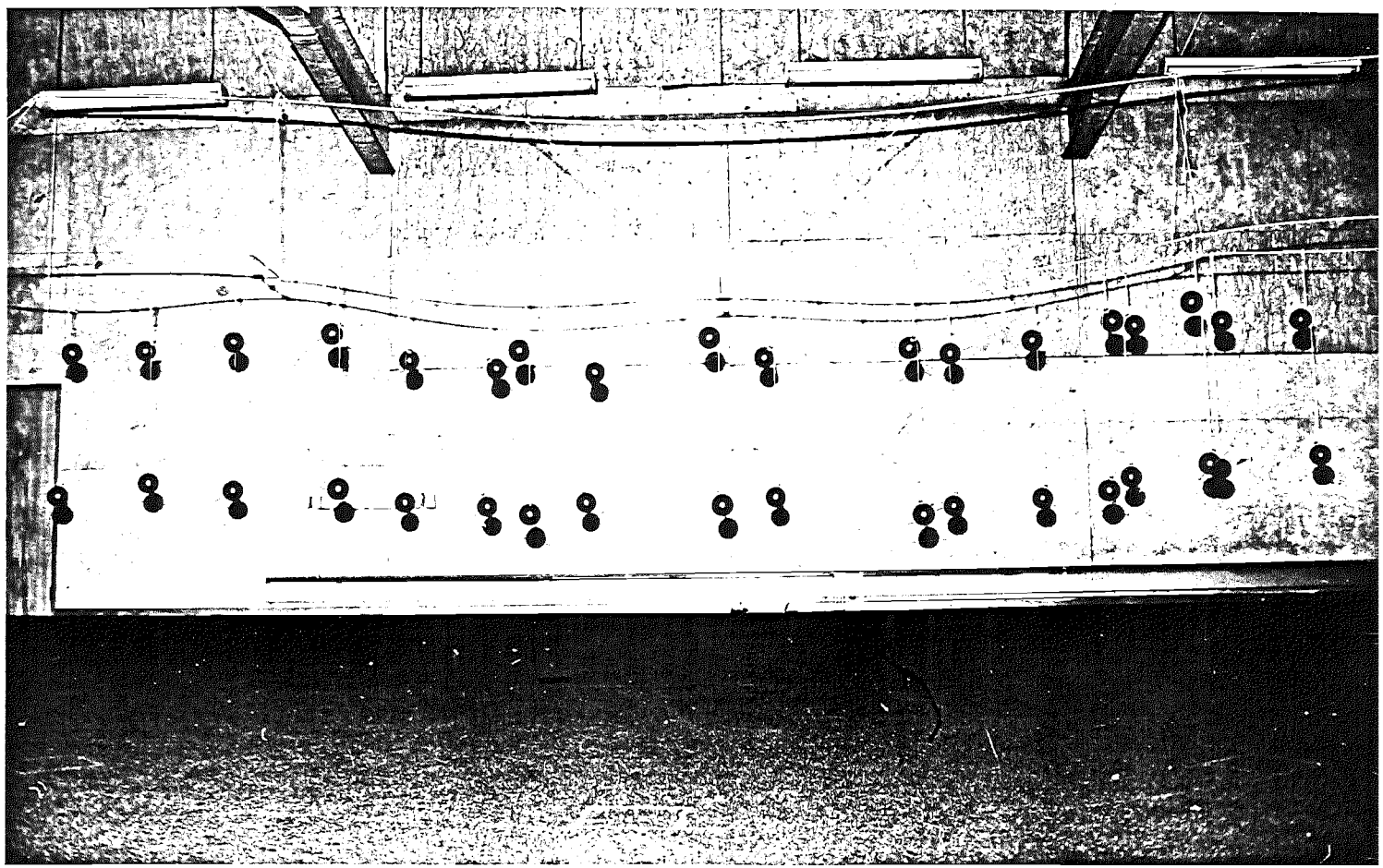
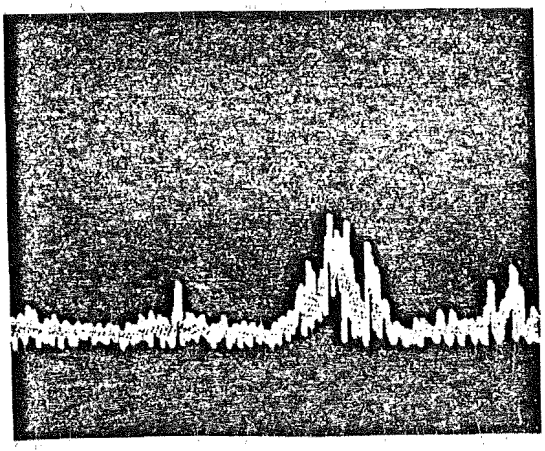
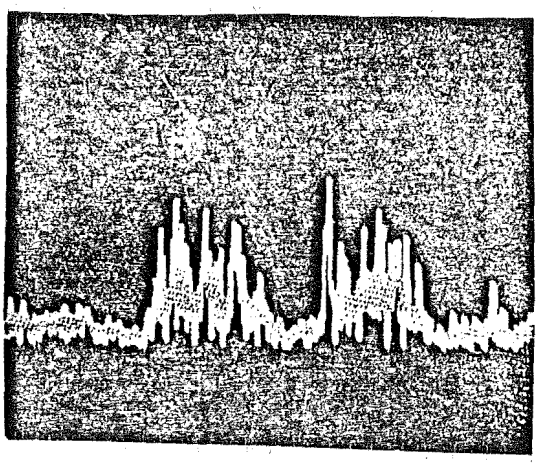


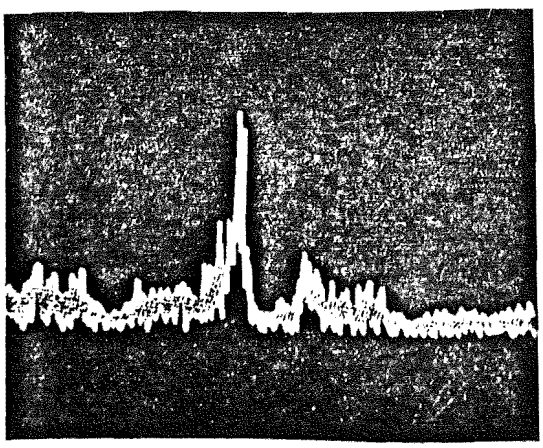
Fig. 20. An artificial shoal target. Forty 11.50 diam. hollow balls are hung on an aluminium frame extending over 10 metres.



a) Signal returned from the artificial shoal at a range of 120 m (a complex tone of center frequency equal to 3200 Hz can be heard and compared with a single tone of the same frequency). Scale = 200 Hz/cm.



b) Signal returned from the artificial shoal at a range of 90 m (a complex tone of center frequency equal to 2400 Hz can be heard and compared with a single tone of the same frequency). Scale = 100 Hz/cm.



c) Signal returned from the 20 in. diam. sphere ar a range of 108 m (audible frequency = 2 900 Hz). Scale = 100 Hz/cm.

Fig. 21. Comparison of echoes given by a sphere and an artificial shoal.

However, since during the time the artificial shoal was being tested - (a) it was not sunk deep enough, and (b) the lake was very rough - therefore as the transducer was elevated to aim at the target at a long distance the sonic beam hit the rough water surface. Strong surface reverberation masked the target echo completely. This also happened to the 20 in. diameter sphere when it was tested in the same conditions. The sphere, at a depth of 21 m under water, could not be detected at a range further than 150 m (it should be remembered that in trial 1 this sphere, at 40 m depth, could be detected from a distance of 600 m when the water surface was just choppy calm). Figure 21(c) shows the signal given by the sphere when detected at the range of 108 m (audible frequency = 2900 Hz). Counting the number of components of the complex signal in Figure 21 (b), and comparing its average amplitude with the amplitude of the signal given by the sphere, it is found that the total power of the complex signal is about 6.16 dB higher than that of the signal given by the sphere. However, since the sphere is further than the artificial shoal, the loss of the echo from the sphere is 3.29 dB higher than that of the echo from the shoal. Hence the target strength of the shoal is about 2.87 dB higher than the target strength of the sphere. Note that, theoretically, the target strength of the shoal is -14.8 dB and target strength of the sphere -18 dB.

4) Recently (February, 1977), during the trials along the coastal area near Whangarei (30 to 40 fathoms of water) it was found that a sonar operator could easily distinguish the sound given by a pinnacle from the sound given by a shoal of fish. The sound returned from a shoal was always wobbling and musical, due to rapid movements of fish (the movements of fish caused both amplitude modulation and frequency modulation in the displayed sound due to rapid variations

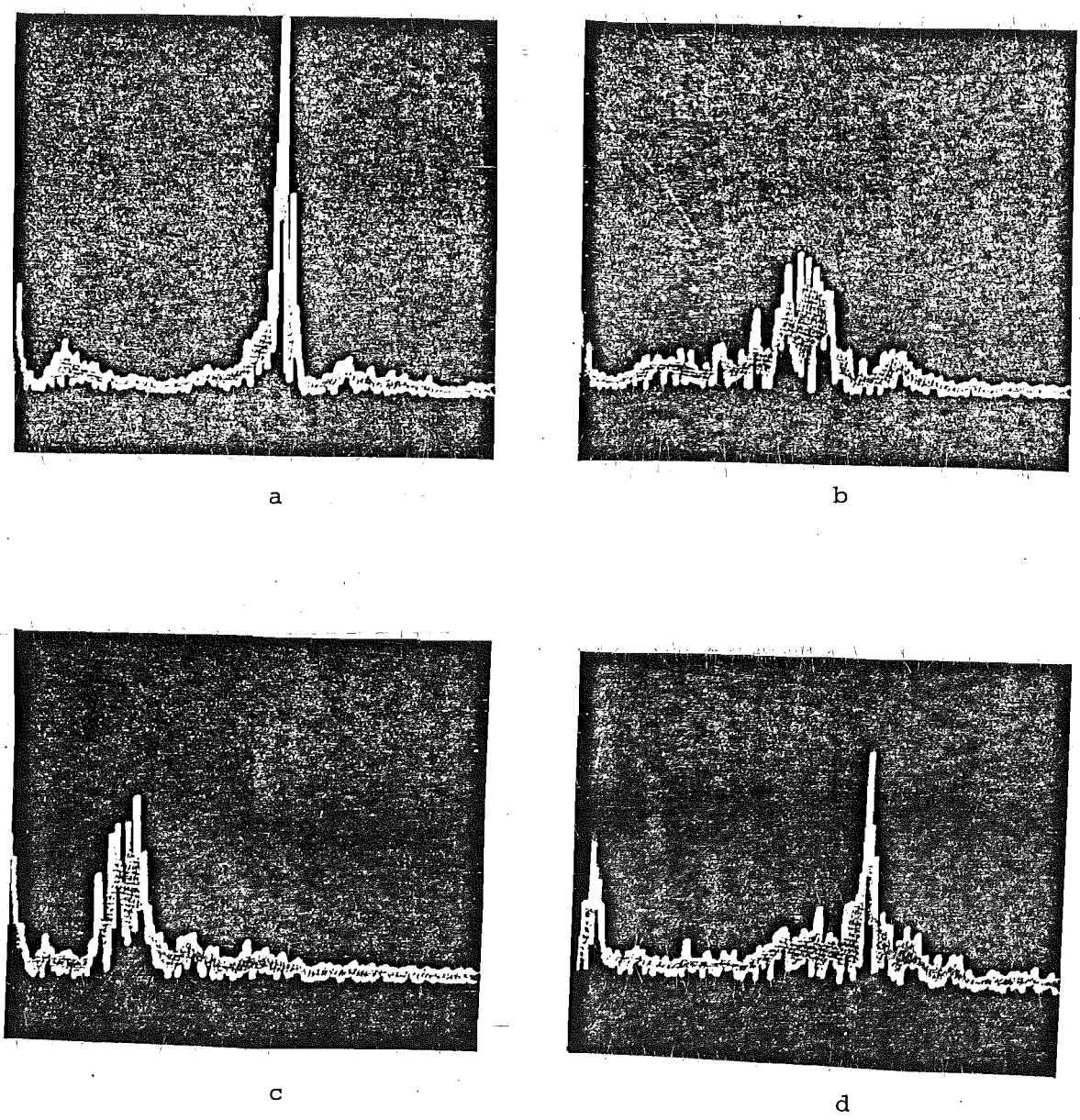
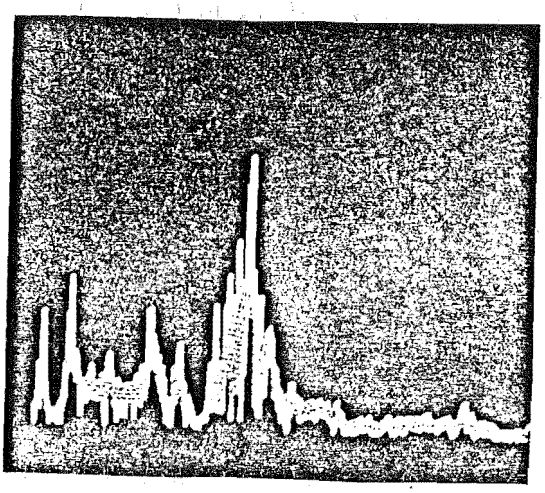
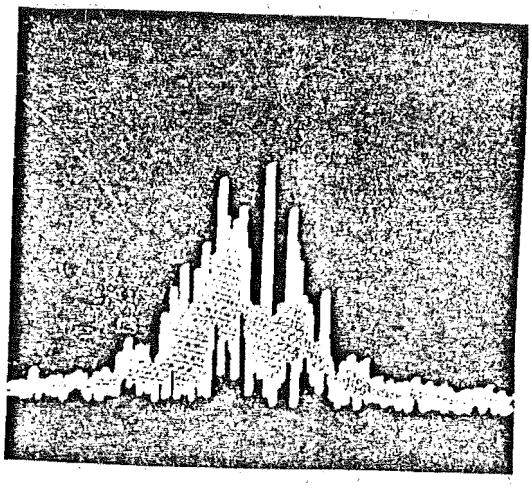


Fig. 22. Comparison of the frequency spectra given by (a) a pinnacle at 555 m (audio frequency = 3700 Hz), (b) a shoal of fish at 450 m (audio frequency = 3000 Hz), (c) a shoal of fish at 255 m (audio frequency = 1700 Hz), and (d) fish around a pinnacle at 570 m (peak frequency = 3800 Hz). The sonar was operated at 8 sec. sweeps, range code = 6.67 Hz/m. Scale on these pictures is 1 kHz/cm. Reference frequency = 0 Hz (first spike).



a) Sound spectrum obtained when approaching a shoal of fish in front of a pinnacle. The sonar was operated at 2 second sweeps, range code = 26.67 Hz/m . The pinnacle is at a range of 98 m (highest peak at 2600 Hz). Scale = 1 kHz/cm. Reference frequency = 0 Hz (first spike).



b) Sound spectrum obtained when detecting a shoal of fish at a range of 495 m (center frequency of this spectrum is 3300 Hz). Scale = 500 Hz/cm. The sonar was operated at 8 sec. sweeps.

Fig. 23. Sound spectra given by two shoals of fish in different situations.

of the back scattering cross-sections of the individuals with respect to the incident angles of the transmitted waves as well as the Doppler shifts of the echoes), while the sound returned from a pinnacle resembled a narrow band, stationary tone. Fish living near rock could be recognized from a long distance of almost 600 m due to the modulation characters introduced into the echo from the rock. The wobbling of the sound became more pronounced as a shoal was approached. Different samples of sound were analysed, and their frequency sepectra are shown in Figures 22 and 23. Frequently, the detected shoals still produced very clear ringing sounds at a range of 500 m. Large individual fish crossing the sonic beam quickly were also frequently detectable.

7.8 CONCLUSION

Methods for use to estimate the average target strength of a shoal of fish, and the reverberation level in a CTFM sonar system, were presented in this chapter. The examples and the experimental results given in the context have elucidated the application of these methods in fish-finding problems. The latest results proved that the binaural sonar can perform the fish-finding task very well at areas having more than 30 fathoms of water. The capability of distinguishing shoals of fish from rocks quickly during the short tracking period is a special character of the CTFM sonar using binaural display. Smith discussed the possibility of improving the detectability of the modulated sound from noise in [11] (the modulation due to the movements of fish). A series of laboratory experiments were carried out to determine the enhancement of the detectability qualitatively when the frequency modulation is presented. These will be described in chapter 8.

REFERENCES

[1] URICK, R.J. Principle of underwater sound. Second Edition, New York, McGraw-Hill Book Company, 1975.

[2] KERR, D.E. Propagation of short radio waves. M.I.T. Radiation Laboratory Series No. 13, McGraw-Hill Book Company, 1951.

[3] HASLETT, R.W.G. Determination of the acoustic back scattering patterns and cross sections of fish. Brit. J. Appl. Phys., Vol.13, 1962, pp.349-357.

[4] STEPHENS, R.W.B. Under water acoustics. Wiley Interscience, 1970, Chapter 5, pp.129-197, by Haslett, R.W.G. Acoustic echoes from targets under water.

[5] HASHIMOTO, T. and MANIWA, Y. Study of the reflection loss of ultransonic wave on fish body by Hilimetre wave. Technical Report, No. 8, Fishing Boat Lab., Ministry of Agriculture and Forestry, Tokyo, Japan, 1956, p.113.

[6] CUSHING, D.H. et al. Measurements of the target strength of fish. J. Brit. I.R.E., Vol.13, 1963, pp.299-303.

[7] N.D.R.C. Division 6, Washington. Physics of sound in the sea. Summary Technical Report, Vol.8, 1946, pp.16, 105, 347, 361 and 463.

[8] MENTZER, J.R. Scattering and diffraction of radio waves. Pergamon Science Series, Pergamon Press Ltd, 1955.

[9] SHISHKOVA, E.V. Study of acoustical characteristics of fish. Modern fishing gear of the world (2), London, Fishing News Book Ltd, 1964, pp.404-409.

[10] KAY, L. An experimental comparison between a pulse and a frequency-modulation echo-ranging system. J. Brit. I.R.E., Vol.20, 1960, pp.785-796.

[11] SMITH, R.P. Transduction and audible display for broad band sonar system. Ph.D. thesis, University of Canterbury, New Zealand, 1972.

[12] FLETCHER, H. Auditory patterns. Reviews of Modern Physics, Vol.12, 1940, pp.47.

[13] BARAKOS, P.A. Under water reverberation as a factor in ASW acoustics. U.S. Navy Underwater Sound Lab. Rep. 620, 1964.

[14] MARSHALL, N.B. Bathypelagic fishes as sound scatterers in the ocean. J. Marine Res., Vol.10, 1951, pp.1-17.

[15] ANDREEVA, T.B. Scattering of sound by air bladders of fish in deep sound scattering ocean layers. Soviet Phys. Acoustics, Vol.10, 1964, pp.17-20.

[16] CHAPMAN, R.P. and MARSHALL, J.R. Reverberation from deep scattering layers in the western North Atlantic. J. Acoust. Soc. Am., Vol.40, 1966, pp.405-411.

[17] STEPHENS, R.W.B. Underwater acoustics. London, Wiley. Interscience, 1970, Chapter 4; Hunter, W.F. An introduction to acoustic exploration, pp.91-127.

[18] URICK, R.J. A sound velocity method for determining the compressibility of finely divided substances. J. Appl. Physics, Vol. 18, 1947, pp.983-

[19] Physics of Sound in the Sea: Part II, Reverberation, Nat. Def. Res. Commn. Div. 6 Sum. Techn. Rep. 8, 1947, p.316.

[20] URICK, R.J. The back scattering of sound from a Hakor bottom. J. Acoust. Soc. Am., Vol.26, 1954, pp.231-235.

[21] MACKENZIE, K.V. Bottom reverberation for 530 and 1030 cps sound in deep water. J. Acoust. Soc. Am., Vol. 33, 1961, pp.1498-1504.

[22] URICK, R.J. and SALING, D.S. Backscattering of explosive sound from the deep sea bed. J. Acoust. Soc. Am., Vol.34, 1962, 1721-1724.

[23] BURSTEIN, Aaron W. and KEANE, J.J. Backscattering of explosive sound from ocean bottoms. J. Acoust. Soc. Am., Vol.36, 1964, pp.1596-1597.

[24] McKINNEY, C.M. and ANDERSON, C.D. Measurements of back-scattering of sound from the ocean bottom. J. Acoust. Soc. Am., Vol.36, 1964, pp.158-163.

[25] ZWICKER, E. Subdivision of the audible frequency range into critical bands (frequenzgruppen). J. Acoust. Soc. Am., Vol.33, 1961, pp.248.

CHAPTER 8

DETECTION OF FREQUENCY MODULATED MULTIPLE COMPONENT TONES IN NOISE

CHAPTER 8

DETECTION OF FREQUENCY MODULATED MULTIPLE COMPONENT TONES IN NOISE

8.1 INTRODUCTION

To predict the masked threshold of a signal in noise, one frequently refers to the critical band theory. Fletcher [1], in 1940, postulated that of all the frequencies in a wide band noise, only those lying within a certain band would produce masking of a pure tone whose frequency lay in the middle of the band. This critical bandwidth is determined by increasing the bandwidth of the masking noise from a few Hz until the masked threshold of the signal does not increase any more. According to Fletcher and Munson [2] and Hawkins and Stevens [3], the critical band may also be defined as the bandwidth of noise whose overall energy is equal to the energy of the pure tone in the center of the band when the tone is just barely masked by the noise. A corrolary of this second definition is that a critical band is the bandwidth of noise whose absolute threshold is equal to the threshold of a pure tone in the center of the band [4]. The critical bandwidths measured according to these three definitions are very close together. Besides, the third definition implies that the energy of a signal contained in a critical band is summated.

Zwicker et al. [5,6] showed that when the bandwidth of a multiple component tone is increased, the loudness remains constant until a critical point is reached, after which the loudness increases. The same effect occurs when the stimulus is a narrow band of noise.

The critical bandwidth at which the loudness summation begins to depend on the spread of energy is approximately the same as that determined by methods involving (i) thresholds, (ii) masking, and (iii) phase [7,8,9]. In the first method (i), the procedure is to find the threshold of a single tone, then repeatedly add, one at a time, a component spaced 10 Hz lower and reduce the amplitude of all individual components equally to find the resulting threshold of the complex tone. There will be a critical point where no further decrease in amplitude of the components occurs. This determines the critical band. In the second method (ii), a narrow band of noise is placed midway between two pure tones of equal amplitude, and the threshold of the noise is repeatedly measured as the two tones are moved further apart in frequency. It is found that the threshold of the noise remains constant until the separation between two tones reaches a critical value, after which it decreases quickly. In the third method (iii), Zwicker [9] measured the just detectable modulation (in terms of the amplitudes of the side bands) of the frequency modulated sound and the amplitude modulated sound, and found that modulation of the latter sound is more easily detectable. However, as the rate of modulation increases and the side bands are spread wider apart, a point is reached beyond which the just detectable modulation is the same for both kinds of sound - the critical band as measured directly by these methods is about two and a half times as wide as the critical band derived by earlier work [1 - 4]. Scharf [10,11] studied the loudness of complex sounds and found the same results as in [5 - 9]. These results were surprisingly corroborated by Hamilton [12] and Greenwood [13] when they used Fletcher's method to measure the critical band. Swets et al. [14] proposed that different assumptions about the shape of the band by the above

mentioned workers resulted in widely varying estimates of the critical bandwidth.

One of the purposes of this chapter is to investigate the role of the critical bands in the summation of power of a complex tone when it is masked by wide band noise. Schafer and Gales [15] measured the detectability of the two, four and eight component tones masked by noise in the frequency region from 600 to 1500 Hz, and showed that the threshold of the complex tones can be improved from 0 to 3 dB even when all the individual components are separated from each other by more than one of Fletcher's critical bands. These results were supported by Green's data [16]. However, Marill [17] used four frequencies of 500, 540, 1060 and 1100 Hz and found that the power summation only occurs for the frequency pair 500 and 540 Hz, not for the pair 500 and 1100 Hz. The data given by these authors [15 - 17] are at low frequencies only. They are not adequate to describe the detectability in the auditory space generated by the binaural sonar.

Further investigation is then given to the case when the multiple component tone is modulated in frequency. It will be shown that the detectability of a frequency modulated multiple component tone can be improved in this dynamic situation to an extent that the critical band theory fails to apply.

8.2 DETECTION OF MULTIPLE COMPONENT TONE IN NOISE (EXPERIMENT 1)

The objective of the experiment is to determine the detectability of multiple component tones masked by noise. The noise used is the pseudo-random noise having a line spectrum of 1 Hz spacing between two successive components. This kind of noise was chosen for the experiment because it resembled the noise due to sea reverberation in the fishing sonar better than the white noise when the spectra of

these noises were compared (see chapter 7). Besides, since the pseudo-random noise has a spectrum of one line per Hz, its power spectrum density (power per Hz) is also the power of each line. So, to simplify the calculation of the noise power in a certain frequency band of bandwidth W Hz (hence there are W lines) it is assumed

$$\begin{array}{c}\text{Noise power in a}\\ \text{frequency band W}\\ \text{(in dB)}\end{array} = \begin{array}{c}\text{Power spectrum density,}\\ \text{or power per line}\\ \text{(in dB)}\end{array} + 10 \log W \qquad (1)$$

The multiple component tones are of 8 Hz spacing between two successive components, and have the bandwidths of 100 Hz (hence 12 components) and 60 Hz (7 components). The center frequencies of these tone groups are 515, 1020, 2020, 3020, 4000 and 4500 Hz.

8.2.1 The Simulation

To produce the signals and the noise used in this experiment three pseudo-random noise generators were specially built. The detailed design is shown in appendix VII. The generator can produce a flat line spectrum in the frequency region from 0 to 50 kHz whose line spacing can be selective among 32, 8, 4, 1 Hz per line, 2, 4, 8, 32 lines per Hz, and almost white noise (2048 lines per Hz). Two generators set at 1 Hz per line were used to simulate a directional noise field by partially cross correlating the two independent noise sources after limiting their frequency spectra from 0 to 5000 Hz. The arrangement to produce this noise field is similar to that in chapter 6 and appendix III.

The third noise generator was set at 8 Hz line spacing and used to generate the multiple component tones. For example, to generate the multiple component tones of 100 Hz bandwidth, the noise source is passed through a low pass filter of 50 Hz bandwidth and then

multiplied with a single tone (carrier) of frequency equal to the center frequency of the desired complex tone. A voltage controlled oscillator was used to generate the carrier, so its frequency (hence the center frequency of the complex tone) could be easily controlled by a voltage function $V(t)$. In the first experiment no frequency modulation is involved, so

$$V(t) = v(o) \qquad (2)$$

The intensity of the complex tone could be varied in twenty steps of 1 dB each and was set near is masked threshold. Figure 1 shows the arrangement to generate the complex tone, and Figure 2 shows the spectrum of a 100 Hz bandwidth tone when analysed on the Tektronix frequency analyser type IL5, at dispersion of 20 Hz/cm and resolution of 10 Hz. Equipments were arranged as in Figure 3.

During all experiments, the noise field was set at a fixed level. The total noise level over the frequency spectrum from 0 to 5000 Hz in each ear was 60 dB (ref. 0.0002 dyne/cm). The 100 Hz noise bands of center frequencies equal to 515, 1020, 2020, 3020, 4000 and 4500 Hz were also measured since the total noise spectrum was not perfectly flat (varying within ± 2 dB) over the frequency range from 0 to 5000 Hz. The detectability of a complex signal was measured by the signal to noise ratio per signal bandwidth (i.e. over the same frequency band of the signal) at its masked threshold. The experimental procedure is as follows.

8.2.2 Procedure

To determine the threshold of a signal masked by noise, the method of limits is one of the most frequenly used methods [18]. The masked threshold is defined by the signal to noise ratio corresponding to the detection of the signal by a subject on 50 per cent of the trials.

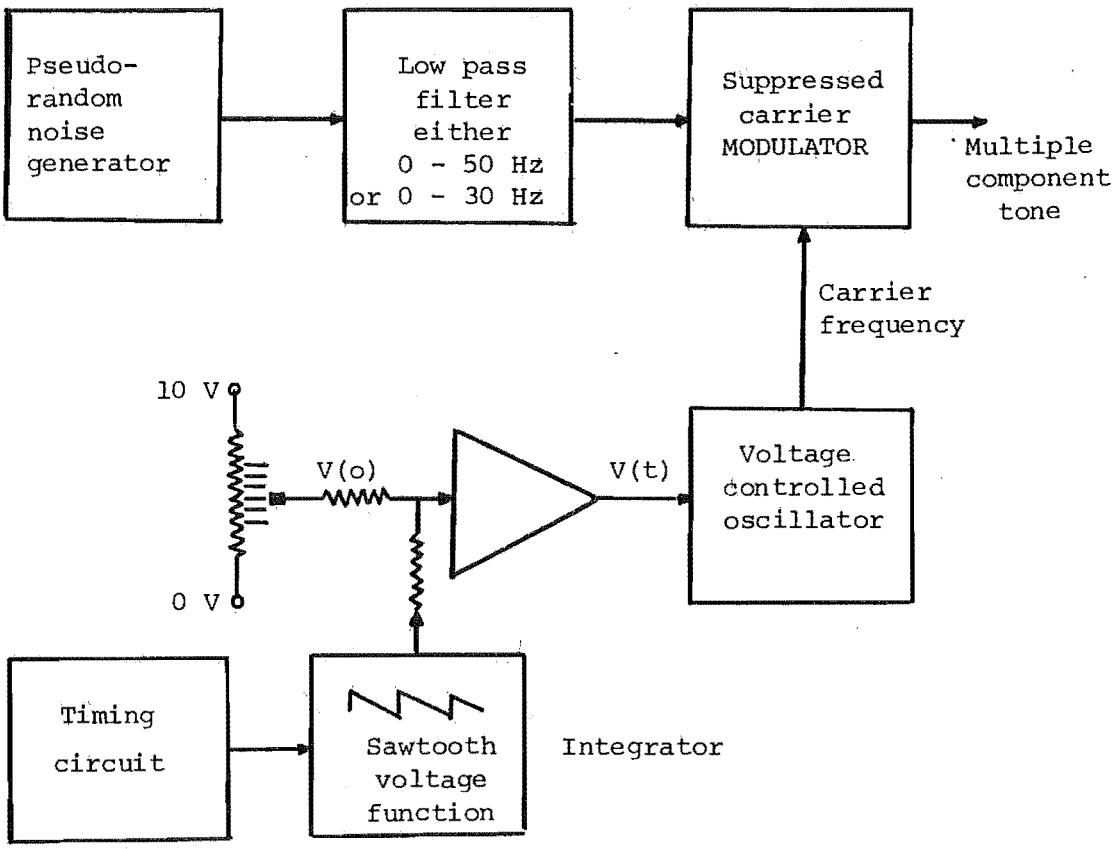


Fig. 1: Arrangement to generate frequency modulated multiple component tone.

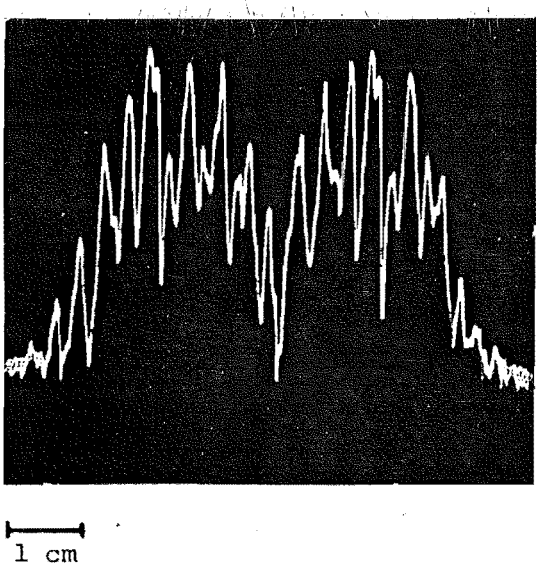


Fig. 2: Spectrum of 100 Hz bandwidth multiple component tone analysed on a frequency analyser.
Scale = 20 Hz/cm,
resolution = 10 Hz.

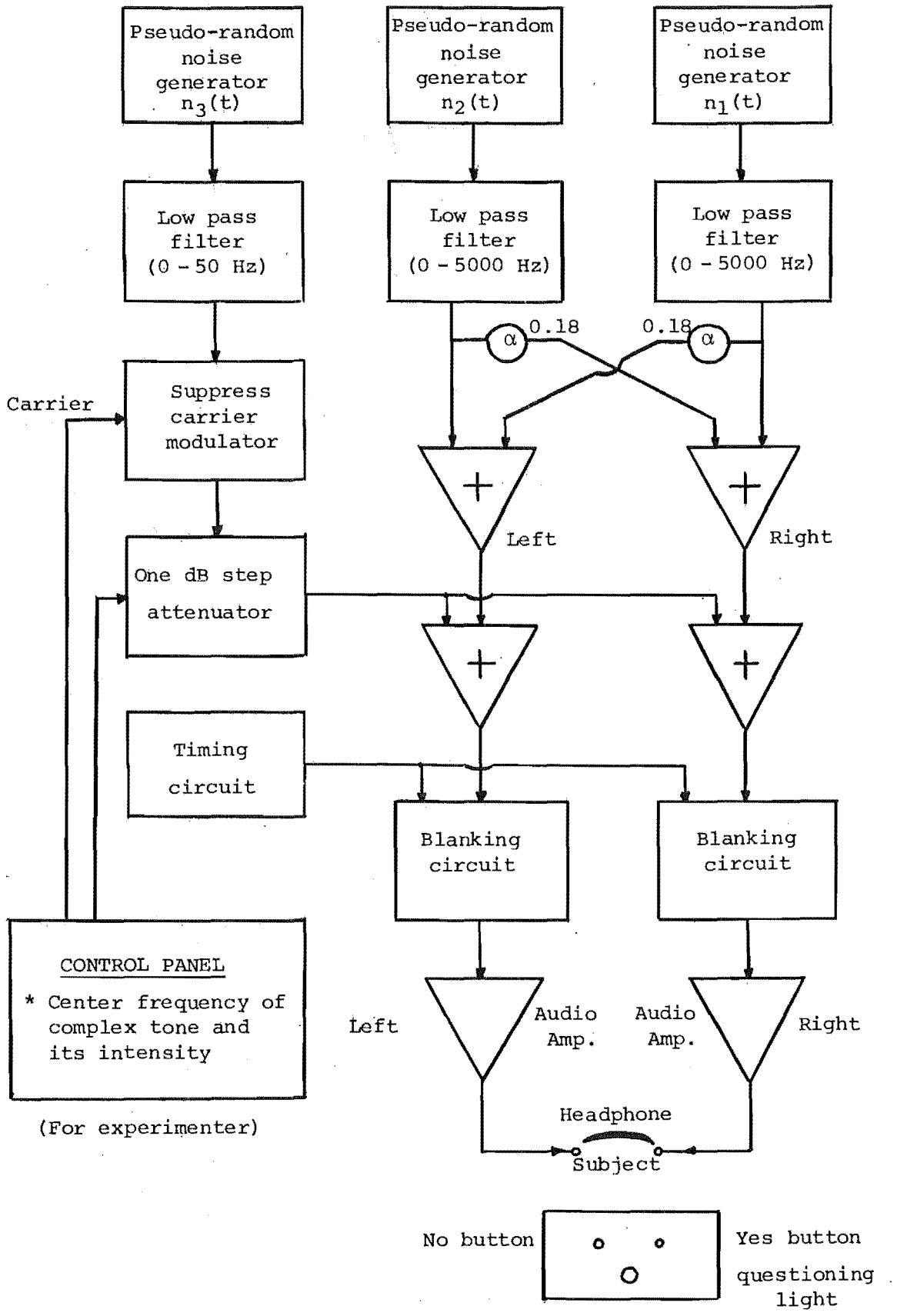


Fig. 3: Arrangement of equipment for the experiments.

Two kinds of stimulus series are generally employed: (1) an ascending series and (2) a descending series. These respectively correspond to increasing and decreasing the signal to noise ratio until the masked threshold is met. Obviously in the descending series a subject knows well how the detected signal sounds while in the ascending series he has to guess or remember (if he has been trained) what the signal is going to be. Therefore the results given in two procedures are different. The experimenter may carry out both procedures then find the average.

During the pre-tests of this experiment it was noticed that the detectability of a signal in noise is also dependent on its transient turn-on. Near the masked threshold it is always more easy to detect a signal which is switched on suddenly from a completely off level ($-\infty$ dB signal to noise ratio) than to detect a signal whose intensity is raised gradually through several steps below its masked threshold, when the experimenter turns the knob of a rotary switch in a natural manner. The latter procedure may result in a masked threshold of about one to two dB higher than that determined by the former procedure. Lengthening the duration of turning the knob does not show any more effect on the detectability.

The purpose of this experiment is to determine the detectability of the complex signal in noise in such a condition that resembles the situation of operating a fishing sonar. Here (in the case of the fishing sonar) the sonar operator is continuously presented with noise due to sea reverberation, and then a complex signal of arbitrary intensity and arbitrary frequency moves into his auditory space. This signal of arbitrary amplitude (due to targets of arbitrary size) is not suddenly switched on since the transducer beam is not idealized. The frequency of the signal is arbitrary since while the sonar is being

operated, the transducer beam must be tilted at a certain angle, hence a deeper target always meets the sonic beam at further range (target range is proportional to the range coding frequency). Thus the experimental procedure was designed as follows.

The subject sat in an anechoic chamber and listened to the noise which was presented to him binaurally. Before the experimenter began to record the results, the subject had a short training period to become familiar with the signals of different intensities and different frequencies which were randomly introduced into the noise field to which he was listening. During the experiment the subject was instructed to listen to the sounds presented continuously in his headphones and was forced to answer the question "Do you hear the signal" by using "Yes" or "No" buttons whenever he saw a bulb light up.

The experimenter's task was to introduce a signal into the subject's headphones by turning a rotary switch to raise the intensity of the signal from $-\infty$ dB (no signal) to a certain level, he then pushed a button to question the subject if he could hear the signal. The frequency and the level of the signal were randomly chosen when the experimenter received the answer from the subject ("Yes" or "No" lights), he made an entry into a table at the tested frequency and S/N ratio - an entry "+" for a "Yes" answer and "-" for a "No" answer - then reset the signal level to $-\infty$ dB before a new trial was commenced. The experimenter could occasionally question the subject if he could hear the signal without introducing any signal into the subject's headphones to verify the subject's judgement. A "No" answer had to be given by the subject in this case every time.

8.2.3 Results

Five subjebts were used in these experiments (S_1, S_2, etc.). Complex signals of center frequencies equal to 515, 1020, 2020, 3020, 4000 and 4500 Hz were presented equally to both ears (no I.A.D ones). The signal bandwidth is 100 Hz in the first series of tests. Since the signal level varied in steps of one dB, it was found that when the signal to noise ratio was near the threshold, the percentage of detection was not always equal to 50% - an increase of one dB in the signal level could change the percentage of detection, say, from 30% to 80%. In this case the higher S/N ratio was chosen to be the masked threshold. The results given by five subjects in the first series of trials are shown in Table 1.

TABLE 1: Masked threshold of complex signal having 100 Hz bandwidth

Signal to Noise Ratio per signal bandwidth at masked threshold (dB)					
Frequency (Hz)	S_1	S_2	S_3	S_4	S_5
515	2	3	2	4	6
1020	1	1	1	2	0
2020	1	1	2	2	0
3020	1.5	1.5	3.5	2.5	0.5
4020	4	2	5	5	4
4500	5.5	4.5	4.5	5.5	5.5

Only subject 5 shows a strange performance - very good detectability in the frequency region from 1000 to 3000 Hz, but poor detectability at 515 Hz. If the critical band of the ear is broader than the bandwidth of the complex signal, and if the power summation hypothesis applies, the critical bandwidth can be determined by

$$\text{Signal to noise ratio per signal bandwidth at masked threshold} = 10 \log \frac{\text{critical bandwidth}}{\text{signal bandwidth}} \quad (3)$$

The average masked threshold of five subjects is converted into the critical bandwidth in Table 2. In this table, the critical bandwidth at the center frequency of 515 Hz is not shown since the signal bandwidth (100 Hz) is larger than the critical bandwidth at this frequency. High masked threshold was obtained, perhaps because the complex signal was spread over more than one critical band.

TABLE 2: The critical bands

Center frequency f (Hz)	Critical bandwidth Δf (Hz)	$\frac{\Delta f}{f}$
1020	126	0.124
2020	132	0.065
3020	155	0.051
4000	251	0.063
4500	324	0.072

To verify this hypothesis, the experiment was repeated with the narrower band signals. Two subjects, S_1 and S_3, were seated for the tests with signals of 60 Hz bandwidth, and these results are compared with their performances in the tests with signals of 100 Hz bandwidth in Table 3.

It is obvious that as the bandwidth of the signal is reduced to 60 Hz, the power of the 515 Hz complex signal can be summated and the detectability is improved. At higher frequencies, the results obtained in the two tests using complex signals of 60 Hz and 100 Hz bandwidths must be considered as very close since the error in the measurements of the masked threshold can be up to ± 1.5 dB. (This includes the error of 1 dB in determining the signal level at the masked threshold, and the error in measuring the power of a noise band.)

TABLE 3: Comparison of the detection-performance of subjects S_1 and S_3 in two different tests using complex signals of 60 Hz and 100 Hz bandwidths.

Frequency	(S/N) per 60 Hz bandwidth at M.T (in dB)			$C.B_2$	(S/N) per 100 Hz bandwidth at M.T (in dB)			$C.B_1$
Hz	S_1	S_3	Average	Hz	S_1	S_3	Average	Hz
515	1.5	1.5	1.5	85	2	2	2	-
1020	2	2	2	95	1	1	1	126
2020	2.5	2.5	2.5	107	1	2	1.5	141
3020	4	4	4	151	1.5	3.5	2.5	178
4000	6.5	5.5	6	239	4	5	4.5	282
4500	8.0	6.0	7	301	5.5	4.5	5	316

$C.B_1$: Critical bandwidth calculated from measuring the masked threshold of 100 Hz bandwidth signals.

$C.B_2$: Critical bandwidth calculated from measuring the masked threshold of 60 Hz bandwidth signals.

A comparison of the critical bands measured in this experiment with the critical bands measured by Hawkins and Stevens [3] and Zwicker [5] is shown in Figure 4. These results are discussed in section 8.5 .

8.3 DETECTION OF FREQUENCY MODULATED MULTIPLE COMPONENT TONE IN NOISE (EXPERIMENT 2)

The purpose of this experiment is to examine the effect of frequency modulation on the detection of the complex tone. Sawtooth frequency modulating function is used to control the center frequency (the carrier frequency) of the multiple component tone. This can partly resemble the Doppler effect in the F.M sonar without making the problem too complicated.

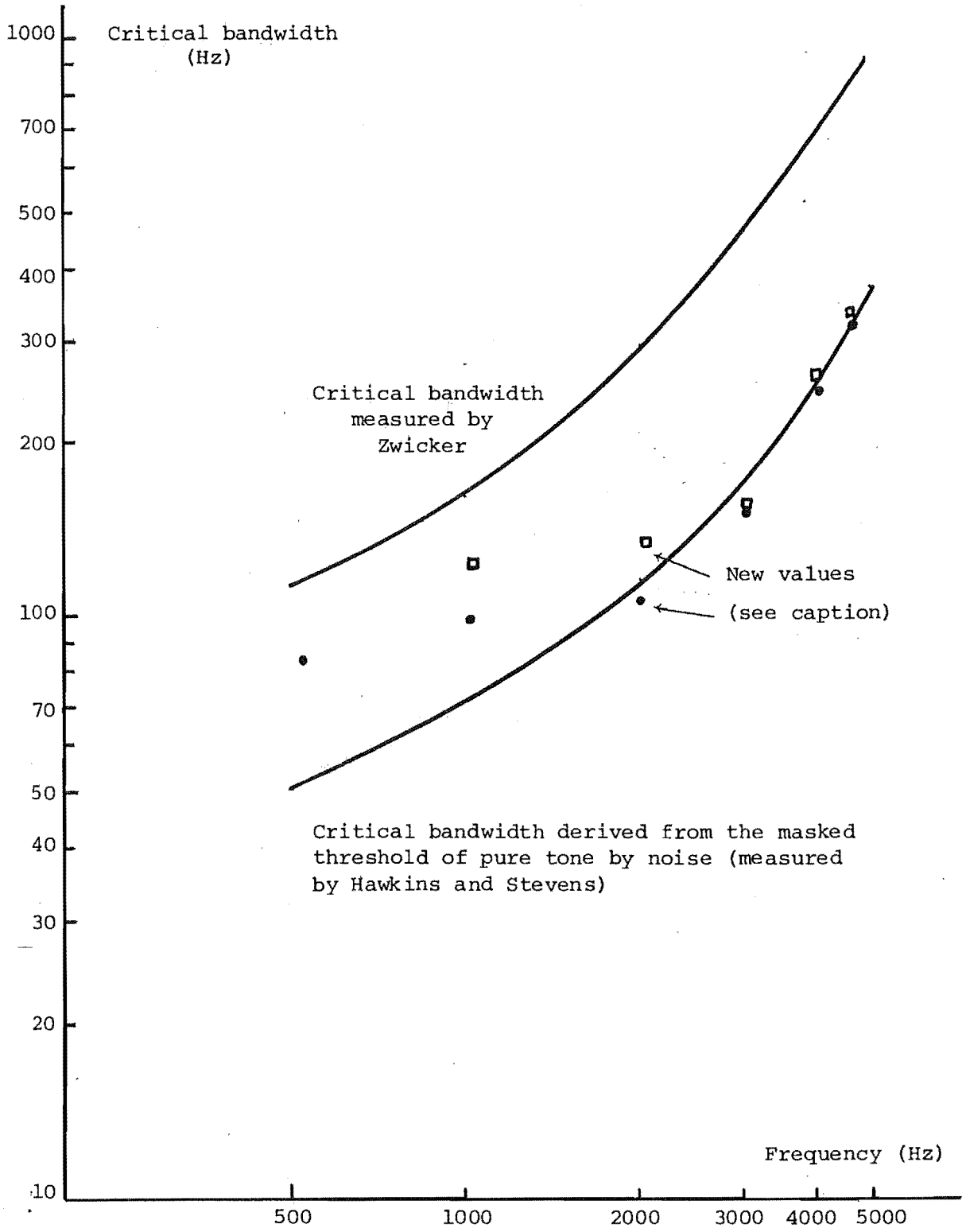


Fig. 4: Comparison of results given by experiment 1 and results given by Zwicker [5] and Hawkins and Stevens [3].

Critical band derived from masked threshold of 100 Hz bandwidth signal,

Critical band derived from masked threshold of 60 Hz bandwidth signal.

Hence the voltage function V(t) in Figure 1 becomes

$$V(t) = v(o) - \alpha t_s \qquad (4)$$

where t_s varies from 0 to 3 seconds then resets during a blanking period of 0.2 sec . Thus, the repetition period, T_s , is equal to 3.2 sec . The constant α is set so that a sawtooth cyclic variation of 100 Hz is obtained. Hence the center frequency of the complex tone is

$$f = f(o) - 33.33\ t_s \quad \text{in Hz} \qquad (5)$$

where f(o) is equal to 515, 1020, 2020, 3020, 4000 and 4500 Hz. Signals of 100 Hz bandwidth are used in this experiment. Using the same experimental procedure, the results given by five subjects, $S_1 - S_5$, are as follows.

TABLE 4: Masked threshold of frequency modulated multiple component tones having bandwidths of 100 Hz. The sawtooth cyclic variation is 100 Hz.

Signal to Noise Ratio per signal bandwidth at masked threshold					
Frequency (Hz)	S_1	S_2	S_3	S_4	S_5
515	2	2	1	3	4
1020	1	1	1	2	0
2020	1	0	0	2	-1
3020	1.5	-0.5	1.5	2.5	0.5
4020	4	2	2	5	3
4500	4.5	3.5	3.5	5.5	3.5

Little improvement on the detection performance is seen on the results given by two subjects, S_1 and S_4 . But significant improvement on the detection performance of the other three subjects, S_2 , S_3 and S_5 , are obvious. The two negative entries - 0.5 and

- 1 dB in Table 4 may mean that at the frequencies of 3020 Hz and 2020 Hz respectively, subjects S_2 and S_5 could detect those F.M signals whose power levels are even lower than the power levels of the noise bands occupied by the signals. However, since the errors in determining the masked thresholds could be up to 1.5 dB, the above results are not unquestionable. On average, in the frequency region from 1000 to 3000 Hz, the masked threshold of these F.M signals is between 0 and 1 dB. The band of noise which effectively masks the signal approximates the bandwidth of the signal of 100 Hz. At 4000 and 4500 Hz, the bands of noise effectively masking the F.M signals are reduced to 209 and 257 Hz respectively. At 515 Hz, although the summation of the signal power is not possible yet, because the signal bandwidth is larger than the critical band, the detection performance of the subjects are improved by up to 2 dB.

8.4 THE EFFECT OF INCREASING THE FREQUENCY MODULATION RATE ON DETECTION PERFORMANCE (EXPERIMENT 3)

We wish to examine the effect of increasing the frequency modulated rate on the detection performance. Let the voltage function $V(t)$ be set at

$$V(t) = v(o) - 2 \alpha t_s \qquad (6)$$

so that the complex tone is modulated at a double rate.

The centre frequencies of the complex signals are

$$f(t) = f(o) - 66.67\ t_s \quad \text{in Hz} \qquad (7)$$

where $f(o)$ is equal to 1020, 2020, 3020, 4000 and 4500 Hz. Since the repetition period is still kept at 3.2 seconds, a sawtooth cyclic variation of 200 Hz is now obtained. The bandwidth of the complex tones is still equal to 100 Hz.

Repeating the experiment with subjects S_2, S_3 and S_5, who show more sensitivity to the frequency modulation in experiment 2, the following results are obtained (Table 5).

TABLE 5: Masked threshold of frequency modulated multiple component tones of 100 Hz bandwidth. The sawtooth cyclic variation is 200 Hz.

Signal to Noise Ratio per signal bandwidth at masked threshold			
Frequency (Hz)	S_2	S_3	S_5
1020	0	1	0
2020	-1	-1	-1
3020	-0.5	1.5	-0.5
4000	1	2	2
4500	2.5	3.5	2.5

At the frequencies of 4000 and 4500 Hz, the detection performance of these subjects is further improved in comparison to the results in experiment 2. The bands of noise effectively masking the 4000 and 4500 Hz signals are now equal to 148 and 193 Hz respectively. At lower frequencies, from 1020 to 3020 Hz, more negative entries appear in Table 5. However these results still do not show sufficient evidence to make any further conclusion about the detection performance of these subjects but that the band of noise which effectively masks an F.M complex signal is, on average, approximately equal to the signal bandwidth (100 Hz in the above tests). This conclusion also implies that if the bandwidth of the F.M complex signal is further reduced, the band of masking noise may be reduced further also. To find evidence for this, the signal bandwidth was reduced to 60 Hz, and the experiment was repeated with subject S_2. It is found that the bands of noise effectively

masking the 1020 and 2020 Hz signals are equal to 60 Hz (see Table 6).

TABLE 6: Masked threshold of the FM signals having bandwidth of 60 Hz, and the bands of effectively masking noise. The sawtooth cyclic variation is 200 Hz.

Frequency Hz	Masked threshold of 60 Hz bandwidth signals dB	Band of masking noise Hz
1020	0	60
2020	0	60
3020	1.5	85
4000	4	151
4500	5.5	213

8.5 DISCUSSION

The critical band derived from the masked threshold of a complex tone measured in experiment 1 is obviously very close to the critical band found by Hawkins and Stevens [3] when measuring the masked threshold of a pure tone. This also agrees with the results given by [1,2,4], and proves that the power of individual components of a complex signal having a bandwidth smaller than the critical band is summated. (According to the results of experiment 1, the power summation of a complex signal is not perfect at the frequencies of 500 and 1000 Hz when the bandwidth of the signal is larger than the critical band.)

The critical band measured in experiment 1 and in [1 - 4] is about 2.5 times narrower than the critical band found by Zwicker and others [5 - 13]. Swet et al. [14] explained that these widely varying estimates of the critical bandwidth are due to different

assumptions on the shape of the critical band, and there is always a non-matching between the noise power measured at the external filter (to produce a narrow band masking noise) and the real amount of noise in the internal filter of a subject (the critical band). However this explanation cannot apply to Hawkins and Stevens' data since they used very wide band noise (100 to 9000 Hz) to mask a pure tone, then deduced the width of the critical band from the masked threshold. For example, if the masked threshold of a pure tone is 20 dB (ratio of signal level to noise level per Hz) then the critical bandwidth is 100 Hz . (A similar method is applied in the experiments described in this chapter.) The reader may argue that in Hawkins and Stevens' experiment, using the formula

$$\text{Signal to Noise Ratio per Hz at masked threshold} = 10 \log \text{Critical bandwidth} \qquad (8)$$

is equivalent to assuming a rectangular shape for the critical filter. However, according to Swets et al.'s data [14] , among four different shapes: single tuned, rectangular, Gaussian (one sigma points), Gaussian (half power points), assumed for the critical band, the rectangular critical band has the largest width. This is contradictory to the fact that Hawkins and Stevens' critical band is 2.5 times smaller than Zwicker's critical band. Swets et al.'s explanation appears to be unsatisfactory.

An increase in the masked threshold of a signal in noise by 4 dB is equivalent to an increase in the critical bandwidth by 2.5 times. The variations in masked threshold of different subjects with normal hearing are sometimes more than 3 dB. The errors in measuring the signal level and noise level are possibly ± 1 dB. The performance of a subject on one day may differ from that on another day by up to 2 dB. Different experimental procedures may

give different results. Perhaps all these factors, together, result in widely varying estimates of the critical bandwidth.

Besides, the masked threshold of the complex signals measured in experiments 1, 2 and 3 are relatively low since the noise used in these experiments is partially correlated and presented binaurally. Several workers [19 - 21] found that a considerable improvement in detectability of signal is obtained when the noises presented at two ears are cross-correlated. Depending on the conditions and the degree of cross-correlation, this improvement in detectability can be raised up to, say, 12 dB. Perhaps this also explains why negative signal to noise ratio per signal bandwidth is obtained in experiments 2 and 3. The improvement in detectability given by the frequency modulation of the signal is further enhanced by the cross correlation character of the masking noise. Further study on the characteristics of the correlated noise is out of the scope of this chapter. Results of experiments 1, 2 and 3 can be summarized as follows:

- In stationary conditions, the masked threshold of a multiple component tone in noise is determined by the critical bandwidth and the signal bandwidth. If the latter is narrower than the former, the power summation law will apply.

- The detectability of a multiple component tone in noise is improved as it is frequency-modulated. This improvement is enhanced as the frequency modulation rate increases and the band of noise effectively masking the multiple component tone is gradually reduced to the bandwidth of the signal. The explanation for the improvement of frequency resolution in the dynamic case (see chapter 4) also applies to this case.

REFERENCES

[1] FLETCHER, H. Auditory patterns. Review of Modern Physics, Vol.12, 1940, pp.47-65.

[2] FLETCHER, H. and MUNSON, W. Relation between masking and loudness. J. Acoust. Soc. Am., Vol.9, 1937, pp.1-9.

[3] HAWKINS, J.E. Jr. and STEVENS, S.S. The masking of pure tones and of speech by white noise. J. Acoust. Soc. Am., Vol.22, 1950, pp.6-13.

[4] HIRSH, I.J. and BOWMAN, W.D. Masking of speech by bands of noise. J. Acoust. Soc. Am., Vol.25, 1953, pp.1175-1180.

[5] ZWICKER, E., FLATTORP, G. and STEVENS, S.S. Critical band in loudness summation. J. Acoust. Soc. Am., Vol.29, 1957, pp.548-557.

[6] ZWICKER, E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am., Vol.33, 1961, pp.248.

[7] GASSLER, G. Ueber dic hörschwelle für schallereignisse mit verschieden breitem frequenzspektrum. Acustica, Vol.4, 1954, pp.408-414.

[8] ZWICKER, E. Dieverdeckung von schmalbandgeranschen durch sinustöne. Acustica, Vol.4, 1954, pp.415-420.

[9] ZWICKER, E. Die grenzen der hörbarkeit der amplituden - modulation und der frequenzmodulation eines tones. Acustica, Vol.3, 1952, pp.125-133.

[10] SCHARF, B. Critical band and the loudness of complex sounds near thresholds. J. Acoust. Soc. Am., Vol.31, 1959, pp.365-390.

[11] SCHARF, B. Loudness of complex sounds as a function of the number of components. J. Acoust. Soc. Am., Vol.31, 1959, pp.783-785.

[12] HAMILTON, P.M. Noise masked thresholds as a function of tonal duration and masking noise bandwidth. J. Acoust. Soc. Am., Vol.29, 1957, pp.507-511.

[13] GREENWOOD, D.D. Auditory masking and the critical band. J. Acoust. Soc. Am., Vol.33, 1961, pp.484-502.

[14] SWETS, J.A. et al. On the width of critical bands. J. Acoust. Soc. Am., Vol.34, 1962, pp.108-113.

[15] SCHAFER, T.H. and Gales, R.S. Auditory masking of multiple tones by random noise. J. Acoust. Soc. Am., Vol.21, 1949, pp.392-398.

[16] GREEN, D.M. Detection of multiple component signals in noise. J. Acoust. Soc. Am., Vol.30, 1958, pp.904-911.

[17] MARILL, T. Detection theory and psychophysics, Tech. Report No. 319, Res. Lab. Electronics, M.I.T., Cambridge, Massachusetts, 1956.

[18] CORSO, J.F. The experimental psychology of sensory behavior. New York: Holt, Rinehart and Winston, Inc., 1967, pp.224-229.

[19] HIRSH, I.J. The influence of interaural phase on interaural summation and inhibition. J. Acoust. Soc. Am., Vol.20, 1948, pp.536-544.

[20] ROBINSON, D.E. and JEFFRESS, L.A. Effect of varying the interaural noise correlation on the detectability of tonal signal. J. Acoust. Soc. Am., Vol.35, 1963,

[21] LANGFORD, T.L. and JEFFRESS, L.A. Effect of noise correlation on binaural signal detection. Vol.36, 1964, pp.1455-1458.

CHAPTER 9

CONCLUSIONS

9.1 SUMMARY OF EXPERIMENTAL RESULTS

Throughout this thesis, several psychophysical experiments using system simulation techniques are described. Step by step investigation of frequency analysis performance in a multiple object auditory space was approached from the very simple situation of two single stationary tones to the extremely complicated situation of two multiple component tones varying with respect to time in a noise contaminated auditory space. The experimental results have explained why a human operator was able to use the new auditory display in realistic conditions with relative ease, as observed in field evaluation, even though the inputs to the ears were very rich in information. This section presents a summary of all principal results obtained from chapter 2 to chapter 8.

1) In chapter 2, it has been shown that the frequency resolution of the ear when two single stationary tones are simultaneously presented is exceedingly poor. Approximately $\frac{\Delta f}{f} \simeq 40\%$, where Δf is the least frequency difference between two tones so that they are perceived as being distinctly separate, and f is the average frequency of two tones. This frequency difference required for the resolution of two tones is about five times larger than Fletcher's critical band, and 2.35 times larger than Zwicker's critical band. When $\frac{\Delta f}{f} < 40\%$, the combination tones subjectively generated by the non-linearity of the ear cannot be suppressed by the cochlear filtering process. Their presence will therefore interfere

with the perception of the fundamentals.

Thus the spatial resolution of the binaural sonar system is very poor when both the system and the objects are in perfectly motionless conditions.

2) However, the range coding frequency is only proportional to distance in perfectly still situations which are rarely obtained. Chapter 3 shows that once normal relative motion between objects and a sonar user takes place the range coding frequency changes to a complex sound pattern uniquely related to the motion. Different patterns of sound given by different kinds of motion (e.g. objects passed aside, head rotation, etc.) were investigated. It is suggested that the binaural sonar users have learned to use these sound patterns, which flow towards the left or the right at different varying rates of change in frequency as objects are passed, rather than the simple range code as described for a static situation.

3) Binaural sound patterns of different rates of change in frequency, resembling the sounds produced by the CTFM sonar, were simulated using an analogue computer (EAI 580) and are described in chapter 4. A new auditory sensation was found and used to define the frequency resolution capability under dynamic conditions. Suppose that at the time $t = t_o$, two time varying tones are initially merged together, then their frequency difference is then gradually increased. The frequency resolution would be defined as the frequency difference between two tones at the instant they "separate" and "slide" in auditory space to their respective positions.

It is shown in chapter 4 that the auditory frequency resolution is significantly improved in the dynamic case. The Doppler effect, involving doubling the rate of change of audio frequency, plays a significant part in this improvement. An increase in relative

velocity of an object may also enhance the resolution capability quantitatively ($\frac{\Delta f}{f}$ may be reduced to about 8%) as well as qualitatively to an extent that signals may be resolved cognitively before the subject's reaction can take place.

4) The perception of audible information given by a finite target extended over a small range and azimuth angle is studied in chapter 5. Each complex tone in the multiple object auditory space is described by four parameters: (i) the centre frequency, (ii) the centre I.A.D., (iii) the band width, and (iv) the extent in I.A.D. of the tone. The most important result found in this study is that discrimination of two complex tones is easier than discrimination of two single tones ($\frac{\Delta f}{f}$ can be reduced by 30% for complex tones), provided that the bandwidths of the tones are restricted to some specific values approximately equal to that of the critical band. Differing from the case of the single tone, where the frequency discrimination does not depend upon the I.A.D's of the tones, the diffusions in I.A.D. of the complex tones worsen the frequency discrimination slightly as the tones are close in direction. As the extents in I.A.D. of the complex tones are increased, the frequency discrimination is worsened and its dependence upon I.A.D. can be observed at larger differences in the directions of two tones. The bandwidth of the complex tone is a good cue for estimating the size of a finite target.

5) Experiments using complex tones are carried out further for the situation of an auditory space contaminated by noise in chapter 6. To simulate a noise contaminated auditory space as similar as possible to that produced by the binaural fishing sonar, two partially correlated noise sources of limited bandwidths (0 - 5000 Hz) are employed. It is found that within the range of the

S/N - SB (signal to noise ratio per signal bandwidth) from 20 dB to 40 dB, the frequency discrimination is about 2 to 4 times worse than that in quiet conditions. This frequency discrimination is gradually improved with increasing of the S/N - SB and may reach the best performance at the S/N - SB of 80 dB. The frequency discrimination is gradually worsened as the S/N - SB decreases till a certain value below which the performance declines very fast and the high tone disappears. The high tone is completely masked at an S/N - SB of 10 dB higher than its masked threshold when it is presented alone, and its unperceivable presence can cause the faint low tone to be shifted in direction.

6) Detection problems in wide beam CTFM sonar are considered in chapter 7. This chapter provides methods for estimating the intensities of the sonic waves reflected from shoals of fish, and volume and bottom sea scatterers. Several examples and results obtained from field experiments are given to elucidate the application of the above methods in fish finding problems. The sounds returned from shoals of fish, large fishes, huge rocks, and the noise due to sea reverberation are also described and shown to be of distinguishable characters. It is suggested that the binaural sonar can perform the fish finding task very well at up to 600 metres range in areas having more than 30 fathoms of water, and the detection performance can be improved due to the movements of fish.

7) A series of laboratory experiments were carried out to compare the detectability of complex tones in stationary and dynamic conditions. In chapter 8, it is shown that in stationary conditions the masked threshold of a complex tone is merely determined by the auditory critical bandwidth and the bandwidth of the tone, provided that the bandwidth of the tone is narrower than a critical band

in order that its power can be summated during the auditory filtering process. However, once the complex tone is modulated in frequency its detectability in noise is improved. This improvement is enhanced as the frequency modulation rate increases, and the band of noise effectively masking the complex tone is reduced to the bandwidth of the tone.

9.2 RECOMMENDATIONS FOR FUTURE RESEARCH

In addition to the above experimental results, theoretical work on the auditory display as well as the sonar system is extensively developed in this thesis. The author recommends that the following parts of this work should be investigated further.

1) From the discussions in chapters 2, 3 and 4, it has been suggested that the cochlea is only the primary frequency selective part of the hearing organ. Frequency is mapped into place on the basilar membrane with a certain amount of distortion (chapter 2). Subjective combination tones are generated once the stimulus consists of many frequency components. The frequency analysis process is carried on further by the collicular neurones. These auditory neurones can produce selective responses to the stimuli which are in accordance with their best frequencies and their own sensitivities in direction, range, rate and functional form of the frequency change. Some neurones can analyse the sound both in frequency and in intensity. Hence in static conditions the frequency resolution of two tones simultaneously presented is poor because the subjective fundamentals and combination tones are only different from each other in frequency, the interference of the combination tones is important. In dynamic conditions the time varying tones produced by the binaural sonar can be different from each other in various characters such as frequency,

rate of modulation, form of modulation, etc., and these differences are further increased for the combination tones. This enables the auditory neurones to discriminate the fundamentals from the combination tone more easily.

A recommendation for future research is to establish a model for the auditory frequency analysis process. This model should be able to explain - (i) the improvement of frequency resolution and detectability in dynamic situations, (ii) the independence of the frequency resolution of two single tones with respect to their I.A.D.'s, (iii) the increasing dependence of the frequency discrimination of two complex tones with respect to their I.A.D's as they become more diffused in direction, and (iv) the significant improvement of frequency discrimination of two complex tones having their bandwidths smaller than the critical band.

2) Section 3.3.2 provides all the mathematical descriptions for the sound patterns produced by a sensory aid when its user rotates his head. It is recommended that an experiment using computor simulation techniques should be carried out to measure the effect of head rotation on angular resolution.

3) In earlier sea trials (1972), Smith suggested that different species of fish could be recognised as "different" by the difference in the character of the echo sounds they produced. Later sea trials by Kay et al. in 1975-77 have not provided adequate data to confirm the above suggestion although they did show that distinguishing between rocks, shoals of fish and individual large fish could be learned quickly by subjects (see chapter 7), and the detection range was greatly improved. The author recommends that laboratory experiments should be carried out to verify Smith's report. Two methods may be used: (i) carrying out more field experiments at

different waters to record the sounds produced by different species of fish, then using these samples of sound in a laboratory experiment to measure the capability of subjects to learn to recognise each of them properly, or (ii) synthesizing different sound patterns having different amounts of amplitude modulation and frequency modulation, then mixing them with the noise due to reverberation recorded in sea trials and measuring the capability of subjects to distinguish these patterns.

4) In Appendix VIII, mathematical treatments for the ambiguity function of the wide band CTFM sonar are given. The resolutions in range and in velocity of the system are defined here, as the differences in range and in velocity of two nearby targets at which the cross correlation between the two echoes is equal to 0.25 of the maximum value obtained when two targets are completely coincident and shown to be very high. In the example of the fishing sonar in Appendix VIII, the resolutions in velocity and in range are found to be $\Delta v = 4 \times 10^{-2}$ ft/s , and $\Delta r = 0.375$ ft. These values correspond to a resolution in range coding frequency of $\Delta f_a \simeq 2$ Hz , and a resolution in rate of change of the range coding frequency equal to $\Delta \dot{f}_a \simeq 0.2$ Hz/s . No physical display system has been shown to provide such high resolution. Thus, investigation of the feasibility to design a real time frequency analyser of high resolution and fast response is recommended as the most essential research for this sonar system.

APPENDIX 1

TABLE 1 - MEAN FREQUENCY RESOLUTION PERFORMED BY SUBJECT S_1 AT DIFFERENT VALUES OF f_{a2}, IAD_1 AND IAD_2. THE SOUND LEVEL = 70 dB.

f_{a2}	IAD_1	IAD_2 (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
1000	0	--	--	--	175	150	190
	4	190	170	165	170	150	165
	8	180	160	175	180	175	185
	12	165	160	150	165	175	160
2000	0	--	--	--	455	455	505
	4	505	440	470	475	480	505
	8	445	465	390	465	505	480
	12	480	465	450	515	470	440
3000	0	--	--	--	450	455	445
	4	470	420	465	455	535	490
	8	565	535	495	490	460	475
	12	535	560	555	545	480	450

TABLE 2 - MEAN FREQUENCY RESOLUTION PERFORMED BY SUBJECT S_2 AT DIFFERENT VALUES OF f_{a2}, IAD_1 and IAD_2. THE SOUND LEVEL = 70 dB.

f_{a2}	IAD_1	IAD_2 (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
1000	0	--	--	--	180	160	200
	4	200	185	200	195	185	200
	8	175	175	180	175	175	160
	12	210	175	170	175	180	185
2000	0	--	--	--	260	265	260
	4	255	260	260	225	300	250
	8	240	265	285	265	260	295
	12	245	215	270	240	260	220
3000	0	--	--	--	270	365	275
	4	275	250	305	330	335	290
	8	265	255	305	290	270	335
	12	335	325	290	285	260	280

TABLE 3 - MEAN FREQUENCY RESOLUTION PERFORMED BY SUBJECT S_3 AT DIFFERENT VALUES OF f_{a2}, IAD_1 and IAD_2. THE SOUND LEVEL = 70 dB.

f_{a2}	IAD_1	IAD_2 (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
1000	0	--	--	--	170	165	180
	4	185	175	175	175	165	175
	8	185	165	170	165	190	190
	12	180	160	165	195	180	180
2000	0	--	--	--	315	355	295
	4	365	380	310	350	350	380
	8	380	325	360	315	335	355
	12	340	355	325	325	350	335
3000	0	--	--	--	350	390	370
	4	300	265	325	370	345	375
	8	390	320	275	355	325	355
	12	390	335	315	305	375	380

APPENDIX II

TABLE 1. MEAN FREQUENCY DISCRIMINATION PERFORMED BY SUBJECT S_1 AT DIFFERENT VALUES OF g_{ac}, IAD_c and JAD_c. THE SOUND LEVEL = 70 dB. THE COMPLEX TONES ARE OF 100 Hz BANDWIDTHS and 2 dB EXTENTS IN I.A.D.

g_{ac}	IAD_c	JAD_c (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
	0	-	-	-	145	130	110
1000	4	100	100	105	150	130	120
	8	100	105	105	135	140	125
	0	-	-	-	195	175	155
2000	4	145	160	180	185	150	150
	8	150	160	180	190	190	160
	0	-	-	-	225	225	215
3000	4	220	215	230	280	245	235
	8	230	210	215	230	250	270

TABLE 2. MEAN FREQUENCY DISCRIMINATION PERFORMED BY SUBJECT S_2 AT DIFFERENT VALUES OF g_{ac}, IAD_c and JAD_c. THE SOUND LEVEL = 70 dB. THE COMPLEX TONES ARE OF 100 Hz BANDWIDTHS and 2 dB EXTENTS IN I.A.D.

g_{ac}	IAD_c	JAD_c (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
	0	-	-	-	135	125	115
1000	4	115	120	125	145	135	120
	8	110	120	115	120	135	140
	0	-	-	-	175	170	165
2000	4	155	160	180	195	175	155
	8	150	165	170	170	180	165
	0	-	-	-	215	195	195
3000	4	180	205	215	260	275	230
	8	170	200	225	215	280	250

TABLE 3. MEAN FREQUENCY DISCRIMINATION PERFORMED BY SUBJECT S_3 AT DIFFERENT VALUES OF g_{ac}, IAD_c and JAD_c. THE SOUND LEVEL = 70 dB. THE COMPLEX TONES ARE OF 100 Hz BANDWIDTHS and 2 dB EXTENTS IN I.A.D.

g_{ac}	IAD_c	JAD_c (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
	0	-	-	-	150	125	125
1000	4	125	135	135	150	145	135
	8	115	125	130	150	160	175
	0	-	-	-	160	160	155
2000	4	150	150	170	190	175	175
	8	155	165	170	190	175	175
	0	-	-	-	230	195	185
3000	4	180	170	230	230	215	185
	8	190	205	215	210	220	220

TABLE 4. MEAN FREQUENCY DISCRIMINATION PERFORMED BY SUBJECT S_2 AT DIFFERENT VALUES OF g_{ac}, IAD_c and JAD_c. THE SOUND LEVEL = 70 dB. THE COMPLEX TONES ARE OF 100 Hz BANDWIDTHS and 4 dB EXTENTS IN I.A.D.

g_{ac}	IAD_c	JAD_c (dB)					
Hz	dB	-10	-6	-2	+2	+6	+10
	0	-	-	-	155	160	140
1000	4	145	155	170	175	165	165
	8	120	125	160	165	175	160
	0	-	-	-	190	185	175
2000	4	165	190	190	205	195	185
	8	155	180	180	190	190	205
	0	-	-	-	340	310	265
3000	4	240	255	295	305	300	275
	8	230	250	270	290	330	295

APPENDIX III

Simulation of a Directional Noise Field

Considering two partially correlated noise sources, $n_1(t) + \alpha n_2(t)$ and $n_2(t) + \alpha n_1(t)$, arranged as in Figure A3.1. The low-pass filters have the same cut-off frequencies of 5000 Hz. $N_1(t)$ and $N_2(t)$ are two independent noise sources set up at the same levels. The coefficient α, varying between 0 and 1, determines the degree of correlation between two noise stimuli.

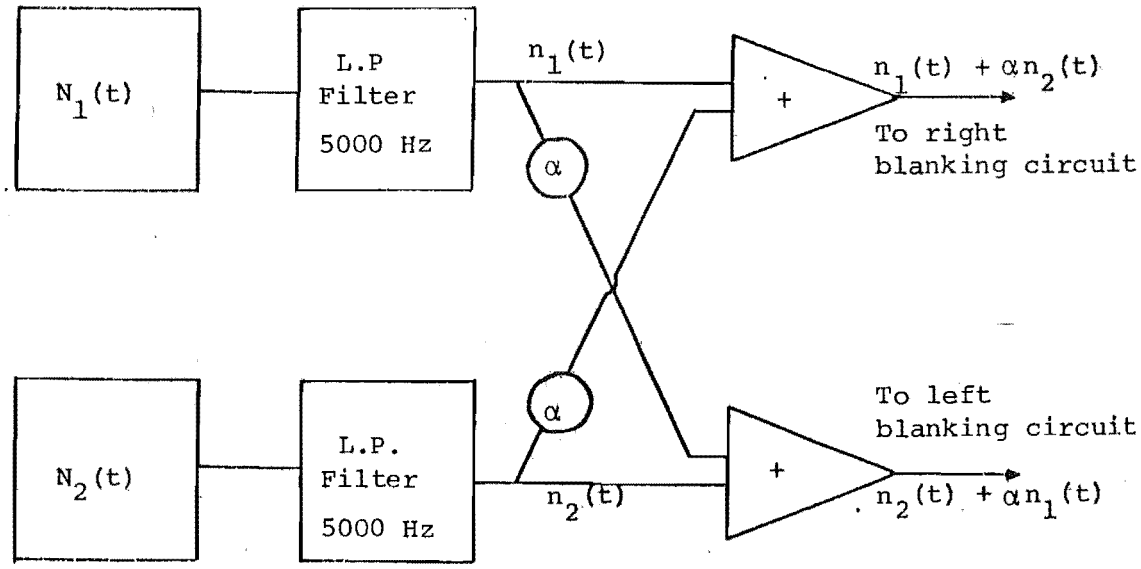


Fig. A3.1. Simulation of a directional noise field.

Since $n_1(t)$ and $n_2(t)$ are two independent noise sources, the phase angles of their frequency components are entirely uncorrelated even though their amplitude spectra are identical. Thus the same frequency $\left(f = \frac{2\pi}{\omega}\right)$ components of two sources can be represented by $A \cos \omega t$ and $A \cos (\omega t + \psi)$ respectively, where ψ is a random variable having an arbitrary value between $-\pi$ and $+\pi$. Using phasor diagram as described in Figure A3.2, it can be shown that the absolute value of the phase difference, ϕ, between two components of the same frequencies, f, of the sources $n_1(t) + \alpha n_2(t)$ and $n_2(t) + \alpha n_1(t)$ is always smaller than the absolute value of ψ.

$$\phi = \psi - 2 \tan^{-1} \frac{\alpha \sin \psi}{1 + \alpha \cos \psi} \qquad \text{(A3.1)}$$

Figure A3.3 shows the variation of ϕ as a function of ψ. Since $n_1(t)$ and $n_2(t)$ are the Gaussian random processes, their phase angles have uniform probability densities. Thus the phase difference ϕ must be evenly distributed. Hence, according to Figure A3.3, about 70% of the value of ϕ will be concentrated within the value $\pm \frac{\pi}{3}$ as α is increased to 0.4. This concentration is increased with increasing value of α.

Actually, the amplitude of the components $A_1 \cos(\omega t + \psi_1)$ and $A_2 \cos(\omega t + \psi_2)$ of the noises $n_1(t)$ and $n_2(t)$ are also the random variables which follow the Rayleigh distribution functions. Hence at a certain value $f = \frac{\omega}{2\pi}$, the direction given by the fusion of the above two components must be indicated by both the I.A.D. and the I.T.D. (interaural time difference). These directional cues vary with respect to frequency and are complicated and non-deterministic (since $\psi = \psi_1 - \psi_2$, A_1 and A_2 are random variables). So it is not necessary to go into details here.

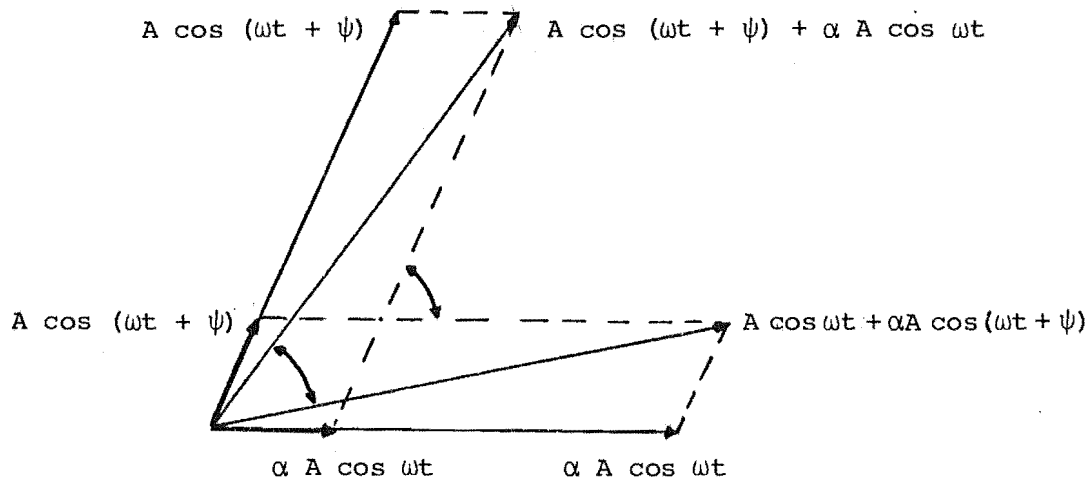


Fig. A3.2. The reduction of phase-difference between two frequency components of the partially correlated noises.

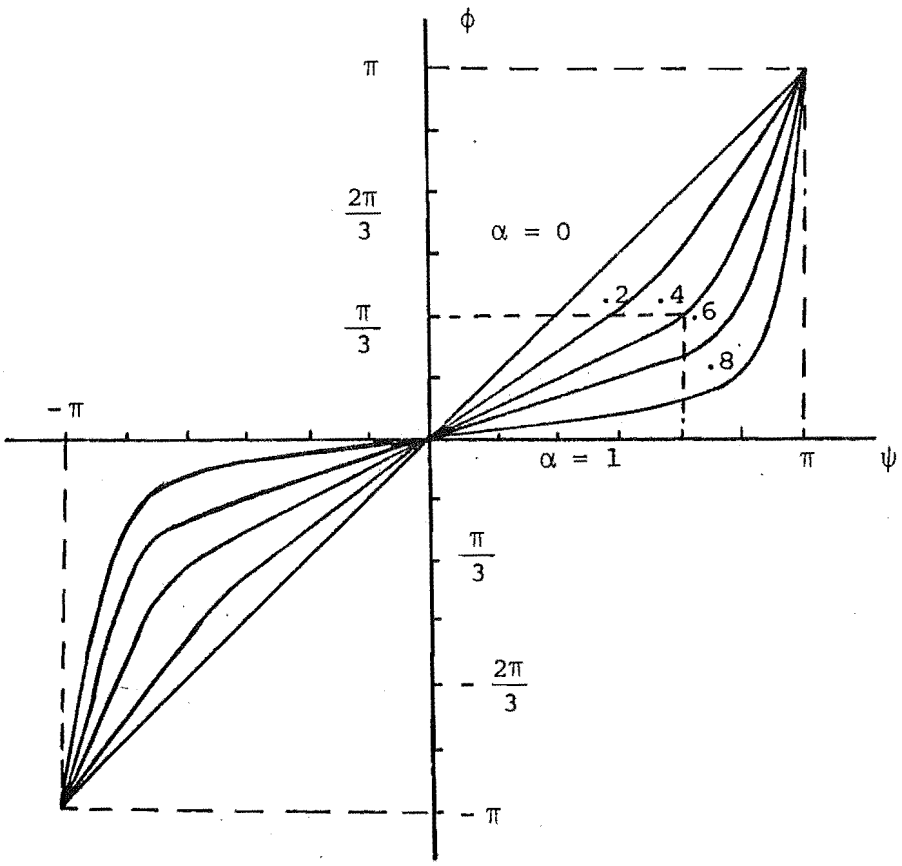


Fig. A3.3. Variation of the phase difference ϕ with respect to ψ .

The above mathematical presentation is adequate to show qualitatively how a directional noise field can be simulated by correlating two noise sources as in Figure A3.1.

APPENDIX IV

TABLE 1. Average Δf_{ac} given by subject S_1 at different values of S/N - S.B , frequency and I.A.D. The total noise level in a 5000 Hz band is 50 dB. The complex tones are of 100 Hz bandwidths and 2 dB extents in I.A.D. Standard deviation is equal to 40, 50 and 85 Hz at g_{ac} of 1000, 2000 and 3000 Hz, respectively.

g_{ac}	(IAD_c , JAD_c)	S/N - S.B (dB)						
Hz	dB	37	32	27	22	17	12	7
	(0 , + 6)	230	260	250	295	340	385	410
1000	(+ 8 , - 2)	225	230	250	265	260	310	310
	(+ 4 , - 10)	220	220	245	255	260	260	280
	(0 , + 6)	455	450	445	490	510*	-	-
2000	(+ 8 , - 2)	390	390	440	450	480*	-	-
	(+ 4 , -10)	380	385	435	430	475*	-	-
	(0 , + 6)	695	740	715	705	1200**	-	-
3000	(+ 8 , - 2)	655	705	685	715	1115**	-	-
	(+ 4 , -10)	630	635	635	625	1040**	-	-

* high tone disappears in 50% of the number of trials,

** high tone is entirely masked.

TABLE 3. Average Δf_{ac} given by subject S_3 at different values of S/N - S.B , frequency and I.A.D. The total noise level in a 5000 Hz band is 50 dB. The complex tones are of 100 Hz bandwidths and 2 dB extents in I.A.D. Standard deviation is equal to 40, 50 and 85 Hz at g_{ac} of 1000, 2000 and 3000 Hz, respectively.

g_{ac} Hz	(IAD_c , JAD_c) dB	S/N - S.B (dB)						
		37	32	27	22	17	12	7
1000	(0 , + 6)	430	430	430	490	500	580	-
	(+8 , - 2)	375	385	370	425	425	505*	-
	(+4 , -10)	365	345	335	345	355	470*	-
2000	(0 , + 6)	540	635	630	630	695*	-	-
	(+8 , - 2)	480	605	610	615	645*	-	0
	(+4 , -10)	485	550	545	585	615*	-	0
3000	(0 , + 6)	745	875	885	1040	1240**	-	-
	(+8 , - 2)	750	750	835	960	1240**	-	-
	(+4 , -10)	655	705	815	920	1235**	-	-

TABLE 2. Average Δf_{ac} given by subject S_2 at different values of S/N - S.B , frequency and I.A.D. The total noise level in a 5000 Hz band is 50 dB. The complex tones are of 100 Hz bandwidths and 2 dB extents in I.A.D. Standard deviation is equal to 40, 50 and 85 Hz at g_{ac} of 1000, 2000 and 3000 Hz, respectively.

g_{ac} Hz	(IAD_c , JAD_c) dB	S/N - S.B (dB)						
		37	32	27	22	17	12	7
1000	(0 , + 6)	275	275	280	290	315	320	365
	(+8 , - 2)	245	255	245	245	255	310	330
	(+4 , -10)	225	245	245	245	245	300	300
2000	(0 , + 6)	535	585	615	715	740*	-	-
	(+8 , - 2)	495	515	625	655	720*	-	-
	(+4 , -10)	500	500	605	610	705*	-	-
3000	(0 , + 6)	795	910	920	1070	1240**	-	-
	(+8 , - 2)	745	910	910	940	1200**	-	-
	(+4 , -10)	750	835	840	910	1165**	-	-

TABLE 1. Target strength and back scattering cross section of some simple forms.

		Symbols	Direction of incidence	Conditions
Any convex surface	$\frac{a_1a_2}{4}$	a_1a_2 = principal radii of curvature r = range $k = 2\pi/\text{wavelength}$	Normal to surface	$ka_1, ka_2 \gg 1$ $r > a$
Sphere Large	$\frac{a^2}{4}$	a = radius of sphere	Any	$ka \gg 1$ $r > a$
Small *	$K\frac{V^2}{\lambda^4}$	V = volume of sphere λ = wavelength	Any	$ka \ll 1$ $kr \gg 1$
Cylinder Infinitely long Thick	$\frac{ar}{2}$	a = radius of cylinder	Normal to axis of cylinder	$ka \gg 1$ $r > a$
Thin	$\frac{9\pi^4a^4}{\lambda^2}r$	a = radius of cylinder	Normal to axis of cylinder	$ka \ll 1$
Finite	$aL^2/2\lambda$	L = length of cylinder a = radius of cylinder	Normal to axis of cylinder	$ka \gg 1$ $r > L^2/\lambda$
	$aL^2/2\lambda(\sin\beta/\beta)^2\cos\theta$	a = radius of cylinder $\beta = kL\sin\theta$	At angle θ with normal	
Plate Infinite (plane surface)	$\frac{r^2}{4}$		Normal to plane	
Finite Any shape	$\left(\frac{A}{\lambda}\right)^2$	A = area of plate L = greatest linear dimension of plate l = smallest linear dimension of plate	Normal to plate	$r > \frac{L^2}{\lambda}$ $kl \gg 1$
Rectangular	$\left(\frac{ab}{\lambda}\right)^2\left(\frac{\sin\beta}{\beta}\right)^2\cos^2\theta$	a, b = side of rectangle $\beta = ka\sin\theta$	At angle θ to normal in plane containing side a	$r > \frac{a^2}{\lambda}$ $kb \gg 1$ $a > b$
Circular	$\left(\frac{\pi a^2}{\lambda}\right)^2\left(\frac{2J_1(\beta)}{\beta}\right)^2\cos^2\theta$	a = radius of plate $\beta = 2ka\sin\theta$	At angle θ to normal	$r > \frac{a^2}{\lambda}$ $ka \gg 1$
Ellipsoid	$\left(\frac{bc}{2a}\right)^2$	a, b, c = semimajor axes of ellipsoid	Parallel to axis of a	$ka, kb, kc \gg 1$ $r \gg a, b, c$
Conical tip	$\left(\frac{\lambda}{8\pi}\right)^2\tan^4\psi\left(1-\frac{\sin^2\theta}{\cos^2\psi}\right)^{-2}$	ψ = half angle of cone	At angle θ with axis of cone	$\theta < \psi$
Average over all aspects Circular disk	$\frac{a^2}{8}$	a = radius of disk	Average over all directions	$ka \gg 1$ $r > \frac{(2a)^2}{\lambda}$
Any smooth convex object	$\frac{S}{16\pi}$	S = total surface area of object	Average over all directions	All dimensions and radii of curvature large compared with λ
Triangular corner reflector	$\frac{L^4}{3\lambda^2}(1-0.00076\theta^2)$	L = length of edge of reflector	At angle θ to axis of symmetry	Dimensions large compared with λ
Any elongated body of revolution	$\frac{16\pi^2V^2}{\lambda^4}$	V = body volume	Along axis of revolution	All dimensions *small* compared to λ
Circular plate	$\left(\frac{4}{3\pi}\right)^2k^4a^4$	a = radius $k = 2\pi/\lambda$	Perpendicular to plate	$ka \ll 1$
Infinite plane strip	$\frac{1}{4\pi k}\left[\frac{\cos\theta\sin(2ka\sin\theta)}{\sin\theta}\right]^2$	$2a$ = width of strip θ = angle to normal	At angle θ	$ka \gg 1$
	$\frac{ka^2}{\pi}$		Perpendicular to strip	$ka \gg 1$ $\theta = 0$

* This table is reproduced from [1] with some modification.

Urick [1], based on reference [3], found that the scattering cross section of a small sphere (Rayleigh scattering region) is:

$$\sigma = 4 \pi K \frac{V^2}{\lambda^4} \qquad \text{(A5.3)}$$

where $K = 61.7$.

Kerr [2], Siegert et al. [8] and Ruck et al. [11] agreed with

$$\sigma = 81 \pi^3 \frac{V^2}{\lambda^4} \qquad \text{(A5.4)}$$

So $K = 20.25 \pi^2$.

This result is quite higher than the result obtained by the U.S. Navy [9] for an arbitrary small object in Rayleigh scattering region:

$$= 16 \pi^3 \frac{V^2}{\lambda^4} \qquad \text{(A5.5)}$$

Figure 1 shows the variation of back scattering cross section of a sphere as its size increases from Rayleigh scattering region to geometrical region. Mentzer [10] determined the variation of scattering cross section of a cylinder with respect to its radius. His result is shown in Figure 2.

Note that the formulas given in Table 1 are directly applicable only if the targets are perfectly rigid. When this condition does not apply, they must be multiplied with the reflection coefficient α_r . Hence, from Table 1

$$T.S = 10 \log F \cdot \alpha_r \qquad \text{(A5.6)}$$

$$\sigma = 4 \pi F \cdot \alpha_r \qquad \text{(A5.7)}$$

where

$$\alpha_r = \beta^2 = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 \qquad \text{(A5.8)}$$

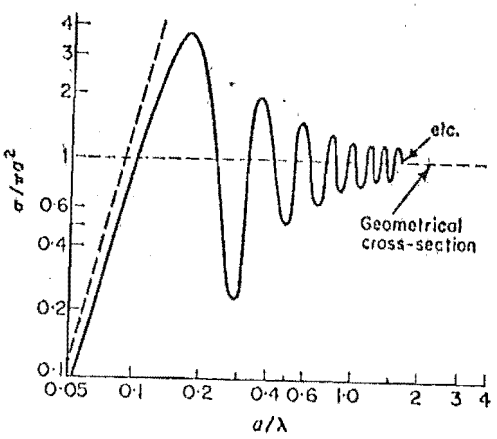


Fig. 1. Ratio of back scattering cross section to geometrical cross section of a perfectly rigid sphere.
a = radius, λ = wave length.
The inclined broken line represents Rayleigh's law,
$\sigma/\pi a^2 = 1.403\ (a/\lambda)^4 \times 10^4$. [2]

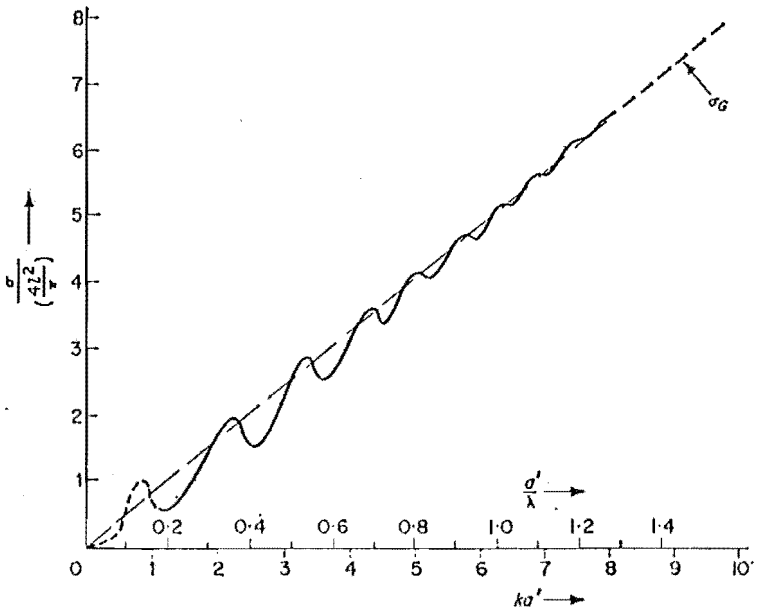


Fig. 2. Back scattering of a short, thick, perfectly rigid cylinder.
a = radius, l = length, $k = 2\pi/\lambda$ = wave number.

Z_2 is the acoustic impedance of the target, Z_1 is the acoustic impedance of the medium, and β is called the reflectivity (see Table 2).

TABLE 2. Acoustic reflectivities of various substances in water

Substance	β (%) in tap water	β (%) in salt water
Air	- 100	- 100 *
Steel	94	94
Brass	92	92
Aluminium	84	83
Granite	83	82
Quartz	82	81
Clay	68	67
Sandstone	68	66
Perspex	35	33
Wet fish bone	26	24
Rubber (pencil eraser)	11	8
Wet fish flesh	4.6	1.9
pc rubber	3	0.32

salt water: salinity 35‰, 15°C

* negative sign corresponds to a change of phase.

REFERENCES

[1] URICK, R.J. Principles of under water sound. Second Edition. New York, McGraw-Hill Book Company, 1975.

[2] KERR, D.E. Propagation of short radio waves. M.I.T. Radiation Laboratory Series, Vol.13, New York, McGraw-Hill Book Company, 1951, pp.445-469.

[3] RAYLEIGH, Lord. Theory of Sound. Vol.2, New York, Dover Publications Inc., 1945, pp.277-311.

[4] PHYSICS OF SOUND IN THE SEA. Part III. Natl. Defense Res. Comm. Div. 6 Sum. Tech. Rept. 8, 1946, pp.358-362.

[5] PROPAGATION OF RADIO WAVES, Committee on Propagation, Natl. Defense Res. Comm. Sum. Techn. Rept. 3, 1946, pp.182.

[6] WILLIS, H.F. Unpublished (British) Report, 1941.

[7] SPENCER, R.C. Back scattering from conducting surfaces. RDB Committee on Electronics, Symposium on Radar Reflection Studies, September, 1950.

[8] SIEGERT, A.J.F. et al. Properties of radar targets. M.I.T. Radiation Laboratory Series. Radar System Engineering. L.N. Ridenowi, Vol.1, New York, McGraw-Hill Book Company, 1947, pp.63-108.

[9] PHYSICS OF SOUND IN THE SEA. Natl. Defense Res. Comm. Div. 6 Sum. Techn. Rept. 8, 1946, pp.463.

[10] MENTZER, J.R. Scattering and Diffraction of Radio Waves. Pergamon Science Series, Pergamon, Oxford, Pergamon Press Ltd, 1955.

[11] RUCK, G.T. et al. Radar cross section hand book. New York, Plenum Press, 1970.

[12] HASLETT, R.W.G. Acoustic echoes from targets under water. Under Water Acoustics, edited by R.W.G. Stephens, London, Wiley-Interscience, 1970.

APPENDIX VI

Calculation of the Equivalent Beam Width of the Projector Hydroplane Combination

From definition, the equivalent beamwidth of the projector hydrophone combination is

$$\Psi = \int_{\theta,\psi} P(\theta,\psi)\ H(\theta,\psi)\ d\Omega \qquad \text{(A6.1)}$$

where $P(\theta,\psi)$ and $H(\theta,\psi)$ are the normalized beam pattern functions of the projector and the hydrophone respectively. θ is the azimuthal angle, and ψ is the elevation. $d\Omega$ is an elemental solid angle. Hence

$$\Psi = \int_{\theta,\psi} P(\theta,\psi)\ H(\theta,\psi)\ \cos\psi\ d\psi\ d\theta \qquad \text{(A6.2)}$$

Ψ is a constant and is independent from the way that the coordinate system is chosen. Let the beam axis be coincident ox .

Let $P_1(\theta)$ and $P_2(\theta)$ be the horizontal beam patterns of the projector and the hydrophone respectively; $P_2(\psi)$ and $H_2(\psi)$ be the vertical beam patterns. Assuming that the latter are idealized and have the same beam widths, Φ .

$$P_2(\psi) = H_2(\psi) = 1 \qquad -\frac{\Phi}{2} < \psi < \frac{\Phi}{2} \qquad \text{(A6.3)}$$

$$= 0 \qquad \text{elsewhere}$$

Then the total equivalent beam width can be approximated by

$$\Psi = 2 \sin\left(\frac{\Phi}{2}\right) \int_0^{2\pi} P_1(\theta)\ H_1(\theta)\ d\theta \qquad \text{(A6.4)}$$

where $$\Theta = \int_0^{2\pi} P_1(\theta)\ H_1(\theta)\ d\theta \qquad \text{(A6.5)}$$

represents the horizontal combination beam width.

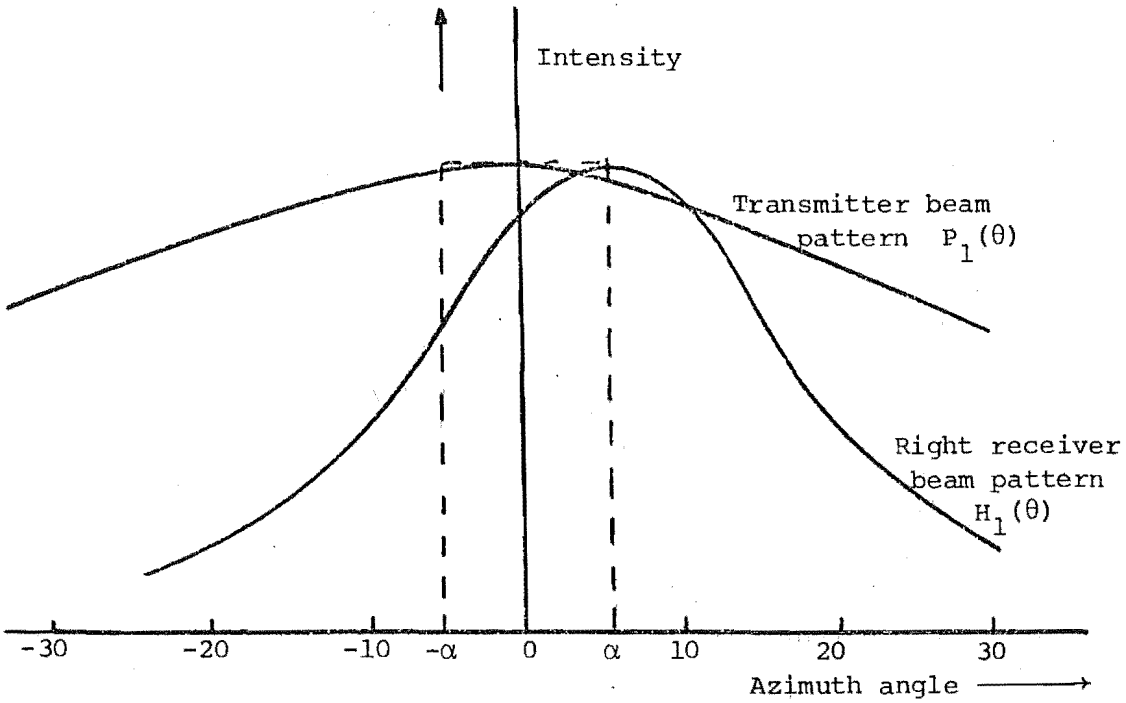


Fig. 1. Horizontal beam patterns of transmitter and receiver in a binaural sonar (left receiver beam not shown).

In this appendix, the combination beam widths of the binaural sonar will be calculated. The ideal voltage responses of the receivers in a binaural sonar have the form

$$A(\theta) = C \exp\left[- \frac{(\theta \pm \alpha)^2}{4 \alpha k} \right] \qquad \text{(A6.6)}$$

where α is the splay angle, k is a constant.

The normalized receiver beam pattern function must be (see Figure 1)

$$H_1(\theta) = \exp\left[- \frac{(\theta \pm \alpha)^2}{2 \alpha k} \right] \qquad \text{(A6.7)}$$

The I.A.D. curve is determined by

$$I.A.D = 20 \log_{10} e . \frac{\theta}{k} \qquad \text{(A6.8)}$$

To obtain a slope of 0.5 dB/degrees for this curve, it requires

$$k = 17.3717^{\circ} = 0.3032 \text{ rad}$$

If the splay angle is $\alpha = 6^{\circ} = \pi/30$, $2 \alpha k = 0.0635 \text{ rad}^2$.

Assuming that the transmitter beam pattern is of bell shape,

$$P_1(\theta) = \exp\left[- \frac{\theta^2}{k_1} \right] \qquad \text{(A6.9)}$$

and has the -3 dB points at $\theta = \pm 30^{\circ}$, then $k_1 = 0.3669 \text{ rad}^2$.

Thus the horizontal combination beam width in the right channel is:

$$\Theta = \int_{-\pi}^{+\pi} \exp\left[- \left(\frac{(\theta - \alpha)^2}{2 \alpha k} + \frac{\theta^2}{k_1} \right) \right] d\theta \qquad \text{(A6.10)}$$

Let $a = 1 + \frac{2 \alpha k}{k_1} = 1.1731$, $b = 2 \alpha = 0.2094$

$c = \alpha^2 = 0.011$, $d = \alpha k = 0.03175$

Rearranging equation (A6.10)

$$= \frac{1}{\sqrt{a}} \exp \frac{C - b^2/4a}{2 d^2} \int_{-\pi\sqrt{a}}^{\pi\sqrt{a}} \exp - \frac{(\sqrt{a}\,\theta - b/2\sqrt{a})^2}{2 d^2} \, d(\sqrt{a}\,\theta)$$

(A6.11)

The above integration has the form of a normal-curve having a mean of $b/2\sqrt{a}$, and a standard deviation equal to d . Since the interval of integration is much greater than the deviation d ($\frac{\sqrt{a}\ \pi}{d} \simeq 19.1$) , the integral is asymmetrically equal to $d\sqrt{2}\,\pi$. Thus

$$\Theta = d\ \frac{2\pi}{a}\ \exp\left\{\frac{C - b^2/4a}{2\ d^2}\right\} \qquad \text{(A6.12)}$$

and the total combination beam width is

$$\Psi = 2\ \sin(\Phi/2)\cdot d\ \frac{2\pi}{a}\ \exp\left\{\frac{C - b^2/4a}{2\ d^2}\right\} \qquad \text{(A 13}$$

Substituting the values of a, b, c, d .

$$\Theta = 0.423 \text{ rad} , \qquad \Psi = 0.1832 \text{ rad}^2$$

or

$$10 \log \Theta = -3.74 \text{ dB} , \qquad 10 \log \Psi = -7.37 \text{ dB}$$

Since the two receiver beams are symmetrical to the axis of the transmitter beam, these values apply to both channels.

APPENDIX VII

DIGITAL NOISE GENERATOR (Pseudo-random Noise Generator)

It is known that digital noise signals can be generated in the form of a sequence of random binary numbers produced by feedback shift registers. The feedback element is an exclusive - OR gate with inputs from the last stage and an intermediate stage which is chosen so that a maximal length sequence can be obtained (Figure 1). The solutions of the feedback function are shown in [1].

A pseudo-random noise generator using N stages of shift register as in Figure 1 has a maximal length sequence of $(2^N - 1)$ states, i.e. after $2^N - 1$ clock pulses the sequence repeats. Since it repeats, it has a line spectrum with a line spacing of $\frac{1}{2^N - 1}$ of the clock frequency. The spectrum is shown in Figure 2.

In the noise generator designed for the experiments in chapter 8, the shift register is 31 stages long. However, to obtain different spacings of 32, 8, 4 and 1 Hz per line, and 2, 4, 8, 32 and 2048 (white noise) lines per Hz, the output stage can be selected as in Table 1. The clock frequency is equal to $f = 1048.575$ kHz.

Figure 3 shows the circuit diagram of the noise generator. A filter having a third order maximally flat (Butterworth) response with a cut-off frequency of 50 kHz and an 18 dB/octave roll off is used in this circuit [2].

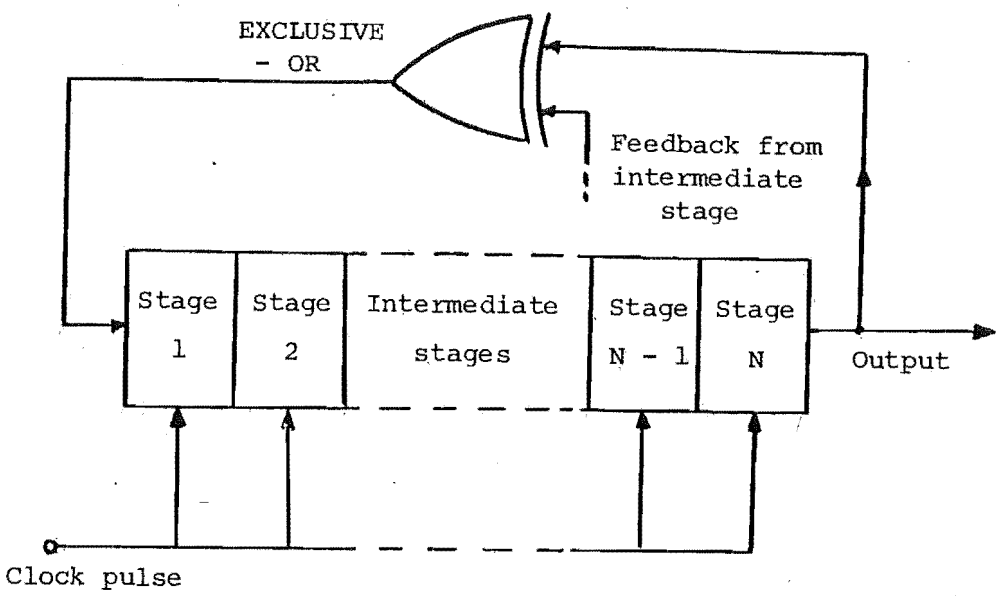


Fig. 1: Principle of a pseudo-random noise generator.

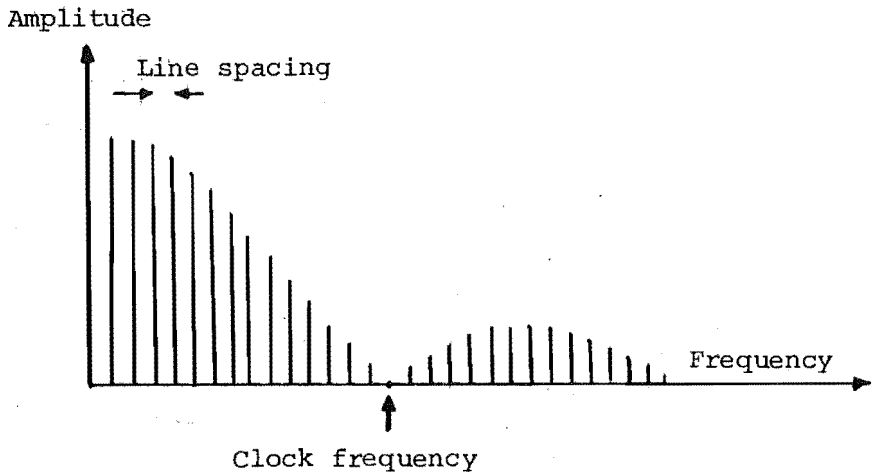


Fig. 2: Spectrum of rectangular waveform of binary sequence generated by a

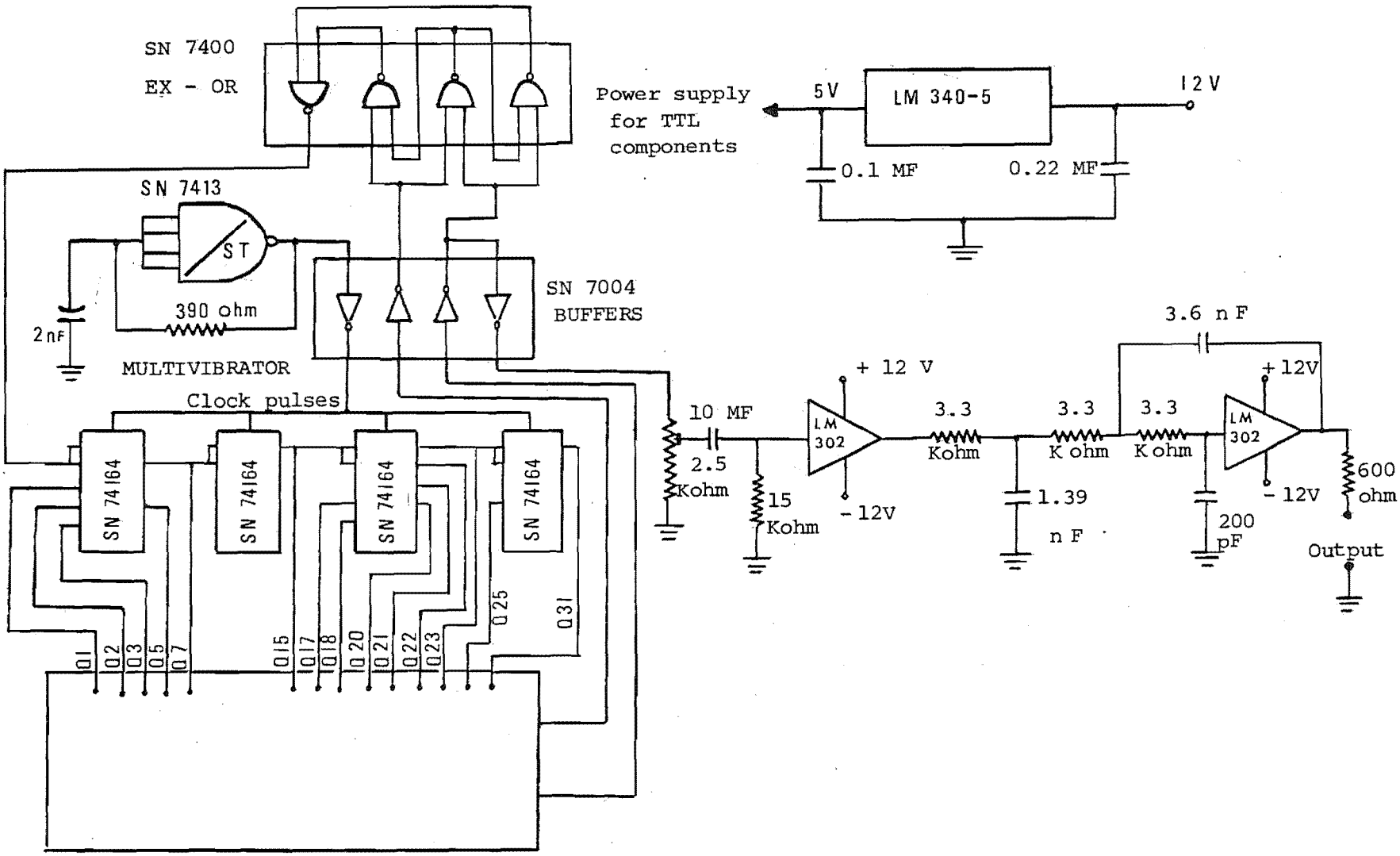


Fig. 3: Circuit diagram of the noise generator.

TABLE 1: Arrangement of the output stage and feedback stage for maximal length sequences, and the corresponding line spacings. Clock frequency = 1048.575 kHz .

Number of stages (output stage)	Feedback tap	Sequence length	Line spacing
15	7	32 767	32 Hz per line
17	5	131 071	8 "
18	7	262 143	4 "
20	3	1 048 575	1 "
21	2	2 097 151	2 line per Hz
22	1	4 194 303	4 "
23	5	8 388 607	8 "
25	7	33 544 431	16 "
31	7	2 147 483 647	2048 "

References

[1] CARLOW, E.F. The logical design of shift counters. Motorola Semiconductor Products Inc., Application Note AN-576, 1972.

[2] BEASTALL, H.R. White noise generator. Wireless World, Vol.78, 1972, pp.127-128.

APPENDIX VIII

VIII.1. AMBIGUITY FUNCTION OF A WIDE BAND CTFM SONAR

Using complex signal notation, the transmitted signal of a CTFM sonar system can be represented by $S_T(t)$.

$$S_T(t) = A \exp \left[2\pi j \left(f_1 t_s + \frac{m}{2} t_s^2 \right) \right] \qquad (1)$$

where $0 \leq t_s \ll T_s$

A is a constant representing the amplitude of this rectangular envelope signal. f_1 is the lower limit frequency, m is the sweep slope, and T_s is the repetition period. At any instant of time t , $t_s = t - t_n$ varies between 0 and T_s , where t_n is the beginning time of the n-th cycle of transmission.

Let C be the sound propagation velocity, and v be the constant approach velocity of a target. The instantaneous range of the target will be

$$r(t) = r_o - v\,t \qquad (2)$$

where r_o is the initial range.

Suppose that the echo signal received at the time t from a target of range r(t) is the replica of the signal transmitted $\tau(t)$ seconds ago, then

$$S_R(t,\tau(t)) = \lambda(t,\tau(t))\, S_T(t - \tau(t)) \,. \qquad (3)$$

where $\lambda(t,\tau(t)) = (1 - \dot{\tau}(t))^{\frac{1}{2}}$ is the normalizing function, and the dot denotes differentiation with respect to time.

If the target velocity is small in comparison to the sound velocity, the instantaneous delay can be approximated by

$$\tau(t) = \tau_o - \alpha\, t \qquad (4)$$

where $\tau_o = \frac{2r_o}{c}$ and $\alpha = \frac{2v}{c}$ are respectively the initial delay and the fractional velocity of the target. Replacing $t = t_n + t_s$ in equation (4)

$$\tau(t) \simeq \tau_n - \alpha \cdot t_s \tag{5}$$

where $\tau_n = \tau_o - \alpha\, t_n$ is the initial delay in the n-th cycle.

If there is now a second target, close to the first one, the second echo signal can be represented by

$$S_R(t,\tau(t) + \Delta\tau(t)) = \lambda(t,\tau(t) + \Delta\tau(t))\; S_T(t - \tau(t) - \Delta\tau(t)) \tag{6}$$

where $$\Delta\tau(t) \simeq \Delta\tau_n - \Delta\alpha \cdot t_s \tag{7}$$

$\Delta\tau_n$ and $\Delta\alpha$ are respectively the differences in time delay and fractional velocity of two targets.

Then the amount of ambiguity on information given by these two targets can be measured by the cross-correlation of the two echo-signals.

$$|\chi(\tau_n,\alpha,\Delta\tau_n,\Delta\alpha)|^2 = \frac{\left| \frac{1}{T_s}\int_o^{T_s} S_R(t,\tau(t))\; S_R^*(t,\tau(t)+\Delta\tau(t))\, dt_s \right|^2}{\left[\frac{1}{T_s}\int_o^{T_s} |S_R(t,\tau(t))|^2 dt_s\right] \cdot \left[\frac{1}{T_s}\int_o^{T_s} |S_R(t,\tau(t)+\Delta\tau(t)|^2 dt_s\right]} \tag{8}$$

The divisor of expression (8) is the product of the powers of the received signals. Substituting the equations (1), (3) and (6) in equation (8), the normalizing functions and the constants, A , cancel each other out after manipulation. The ambiguity function becomes

$$|\chi(\tau_n,\alpha,\Delta\tau_n,\Delta\alpha)|^2 = \left|\frac{1}{T_s}\int_o^{T_s} \exp\, 2\, j\; Z(t)\;\; dt_s\right|^2 \tag{9}$$

where

$$Z(t) = f_1\Delta\tau(t) + m t_s \Delta\tau(t) \;\; - m\tau(t)\Delta\tau(t) - \frac{m}{2}(\Delta\tau(t))^2 \tag{10}$$

These above expressions, (9) and (10), show that the amount of ambiguity varies within the (τ_n,α) space. If the ambiguity contours, defined by $|\chi(\tau_n,\alpha,\Delta\tau_n,\Delta\alpha)|^2$ = constant, are plotted in the (τ_n,α) plane, and the $(\Delta\tau_n,\Delta\alpha)$ sub-planes, they may vary as in Figure 1. Substituting equations (5) and (7) in equation (10), we have

$$z(t) = [f_1 - m\tau_n - \frac{m}{2}\Delta\tau_n]\Delta\tau_n + [m\Delta\tau_n(1+\alpha+\Delta\alpha) - (f_1 - m\tau_n)\Delta\alpha]t_s - [m\Delta\alpha(1+\alpha+\frac{\Delta\alpha}{2})]t_s^2 \quad (11)$$

If the target velocity is assumed to be much smaller than the sound velocity (see chapter 3) ($\Delta\alpha < \alpha << 1$), and the period of repetition T_s , and the sweep slope m are chosen so that the maximum delay given by the furthest target is much smaller than T_s and $f_1 >> m\tau_n$, the ambiguity function can be approximated by

$$|\chi(\Delta\tau_n,\Delta\alpha)|^2 \simeq |\frac{1}{T_s}\int_o^{T_s} \exp\{-2\pi j[m\,\Delta\alpha\, t_s^2 - (m\,\Delta\tau_n - f_1\Delta\alpha)t_s]\}dt_s|^2 \quad (12)$$

τ_n and α are cancelled out in expression (12). This means that the variation of the ambiguity function in the (τ_n,α) plane is very small. $|\chi|^2$, now, depends on the differences in delay and in velocity of the targets only.

Let $$U^2 = m\,\Delta\alpha \quad (13)$$

$$V = \frac{m\,\Delta\tau_n - f_1\,\Delta\alpha}{\sqrt{m\,\Delta\alpha}} \quad (14)$$

Expression (12) can be rearranged as follows:

$$|\chi(\Delta\tau_n,\Delta\alpha)|^2 = |\frac{1}{2\,U\,T_s}\int_{V-2\,U\,T_s}^{V} \exp\{-j\frac{\pi}{2}x^2\}\,dx|^2 \quad (15)$$

where x is an arbitrary variable.

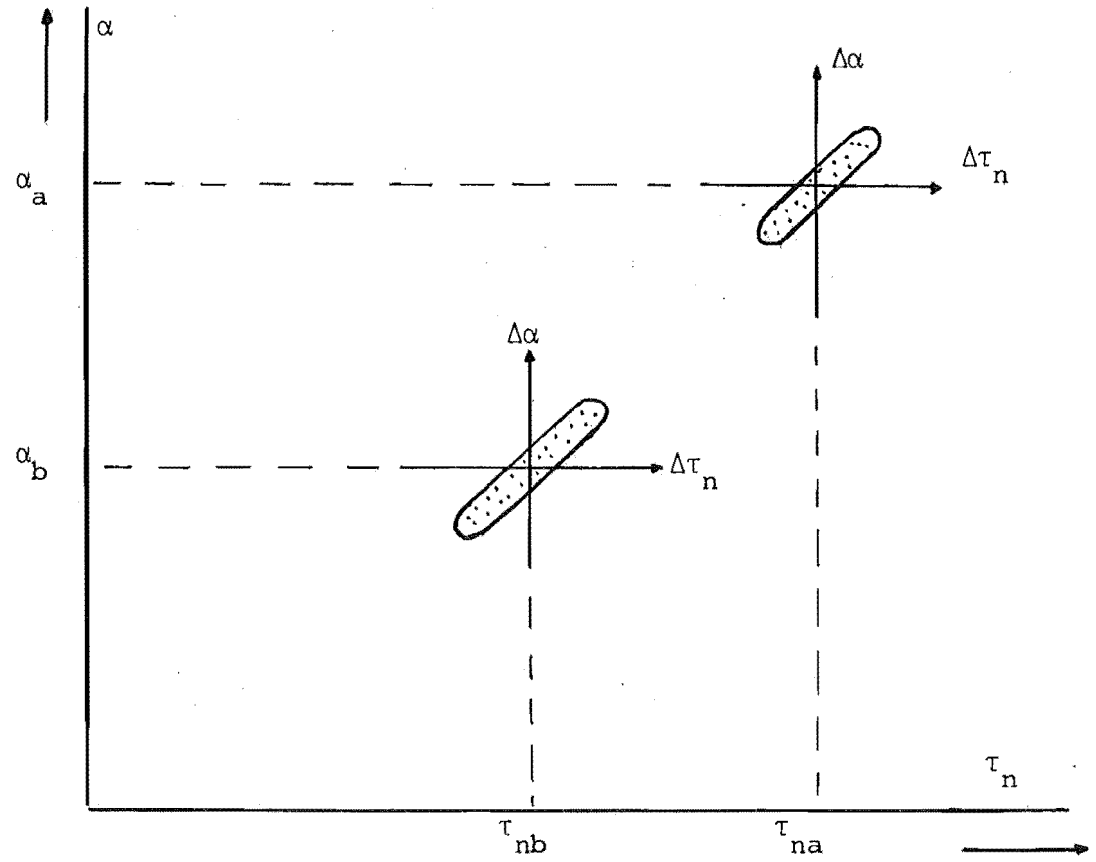


Fig. 1: Variation of the ambiguity contour in the (τ_n, α) plane.

Using the Fresnel's integrals defined by

$$c(z) = \int_0^z \cos\left(\frac{\pi}{2} x^2\right) dx \quad \text{and} \quad s(z) = \int_0^z \sin\left(\frac{\pi}{2} x^2\right) dx$$

The ambiguity function can be expressed as follows

$$\left|\chi(\Delta\tau_n, \Delta\alpha)\right|^2 = \frac{1}{4 U^2 T_s^2} \left[\left(C(V) - C(V - 2UT_s)\right)^2 + \left(S(V) - S(V - 2UT_s)\right)^2\right] \tag{16}$$

or

$$\left|\chi(\Delta\tau_n, \Delta\alpha)\right|^2 = \frac{1}{4 \Delta\alpha W T_s} \left[\left(C(V) - C(V - 2UT_s)\right)^2 + \left(S(V) - S(V - 2UT_s)\right)^2\right] \tag{17}$$

Observing the characteristics of the Fresnel's integrals (Figure 2), C(z) and S(z) are limited by 0.8 and tends to 0.5 when z is large enough, say, greater than 5 . If $\Delta\alpha$ is kept constant, $\left|\chi(\Delta\tau_n, \Delta\alpha)\right|^2$ can be plotted as a function of $\Delta\tau_n$. The following example shows that $\left|\chi(\Delta\tau_n)\right|^2_{\Delta\alpha}$ has the form of a gating function where the mid-value is determined by

$$\left|\chi(\Delta\tau_n, \Delta\alpha)\right|^2_{mid} = \frac{1}{2 \Delta\alpha W T_s} \tag{18}$$

Thus, the ambiguity function of an CTFM system is mainly dominated by the product of the signal bandwidth and its period of repetition. It decreases with increasing this product.

Example

Considering the following example of an underwater sonar for use in fishing. The signal parameters are as follows: f_1 = 40 kHz, w = 40 kHz, T_s = 3.2 secs, m = 12.5 kHz/s. (These parameters are chosen so that a typical range of 200 yards will be coded by a beat note of 3000 Hz.) $\left|\chi(\Delta\tau_n, \Delta\alpha)\right|^2$ will be plotted as a function of $\Delta\tau_n$ for different values of $\Delta\alpha$.

a) $\Delta\alpha = 0$, then $\Delta\tau_n = \Delta\tau$.

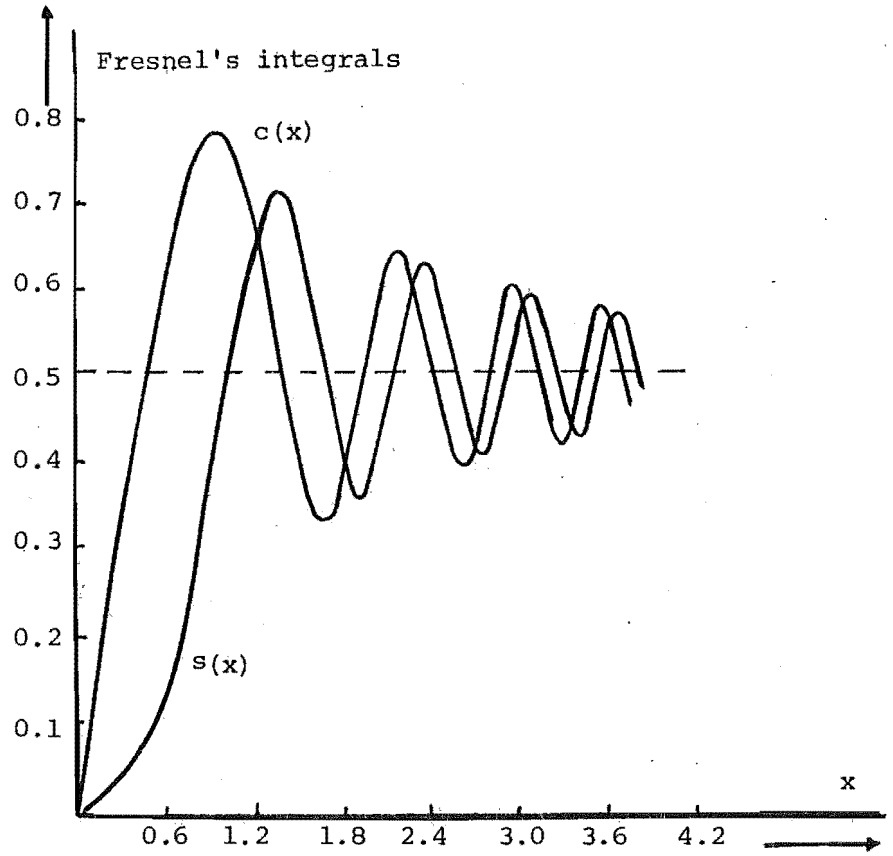


Fig. 2: Characteristics of the Fresnel's integrals.

$$|\chi(\Delta\tau_n, 0)|^2 = \left|\frac{\sin \pi W \Delta\tau_o}{\pi W \Delta \tau_o}\right|^2 = \left|\frac{\sin \pi W \Delta\tau_n}{\pi W \Delta\tau_n}\right|^2 \qquad (19)$$

This function decreases very fast from 1 as $\Delta\tau_n$ is increased. The first zero is $\Delta\tau_n = 2.5 \times 10^{-5}$ s . The eleventh peak of magnitude 9.1×10^{-4} is found at $\Delta\tau_n = 2.625 \times 10^{-4}$ s.

b) As soon as $\Delta\alpha$ reaches 10^{-4} , $2UT_s \simeq 7.2$ and is very large already. Therefore

(i) For $0 \leq \Delta\tau_n \leq \frac{\Delta\alpha\, f_1}{m}$, then $\frac{-f_1 \Delta\alpha}{\sqrt{\Delta\alpha}} \leq V \leq 0$; and $(V - 2UT_s)$ is very negatively large. Hence

$$|\chi(\Delta\tau_n, \Delta\alpha)|^2 \doteq \frac{1}{4 \Delta\alpha W T_s} \left[(0.5 + C(V))^2 + (0.5 + S(V))^2\right] \qquad (20)$$

(ii) For $\Delta\tau_n \geq \frac{\Delta\alpha}{m}(f_1 + 2mT_s)$; then $V \geq 2UT_s$ and is very large

$$|\chi(\Delta\tau_n, \Delta\alpha)|^2 \doteq \frac{1}{4 \Delta\alpha W T_s} \left[(0.5 - C(V - 2UT_s))^2 + (0.5 - S(V - 2UT_s))^2\right] \qquad (21)$$

(iii) For $\frac{\Delta\alpha\, f_1}{m} < \Delta\tau_n < \frac{\Delta\alpha}{m}(f_1 + 2mT_s)$, then $0 < V < 2UT_s$;

Rearranging equation (17)

$$|\chi(\Delta\tau_n, \Delta\alpha)|^2 \doteq \frac{1}{4 \Delta\alpha W T_s} \left[(C(V) + C(2UT_s - V))^2 + (S(V) + S(2UT_s - V))^2\right] \qquad (22)$$

From Figure 3 (a - c), it could be seen that the ambiguity function is mainly raised in this region and has a gating form with the mid-value determined by

$$|\chi(\Delta\tau_n, \Delta\alpha)|^2_{mid} \doteq \frac{1}{2 \Delta\alpha W T_s} \qquad (23)$$

Obviously this is inversely proportional to the product of the signal bandwidth and its period of repetition.

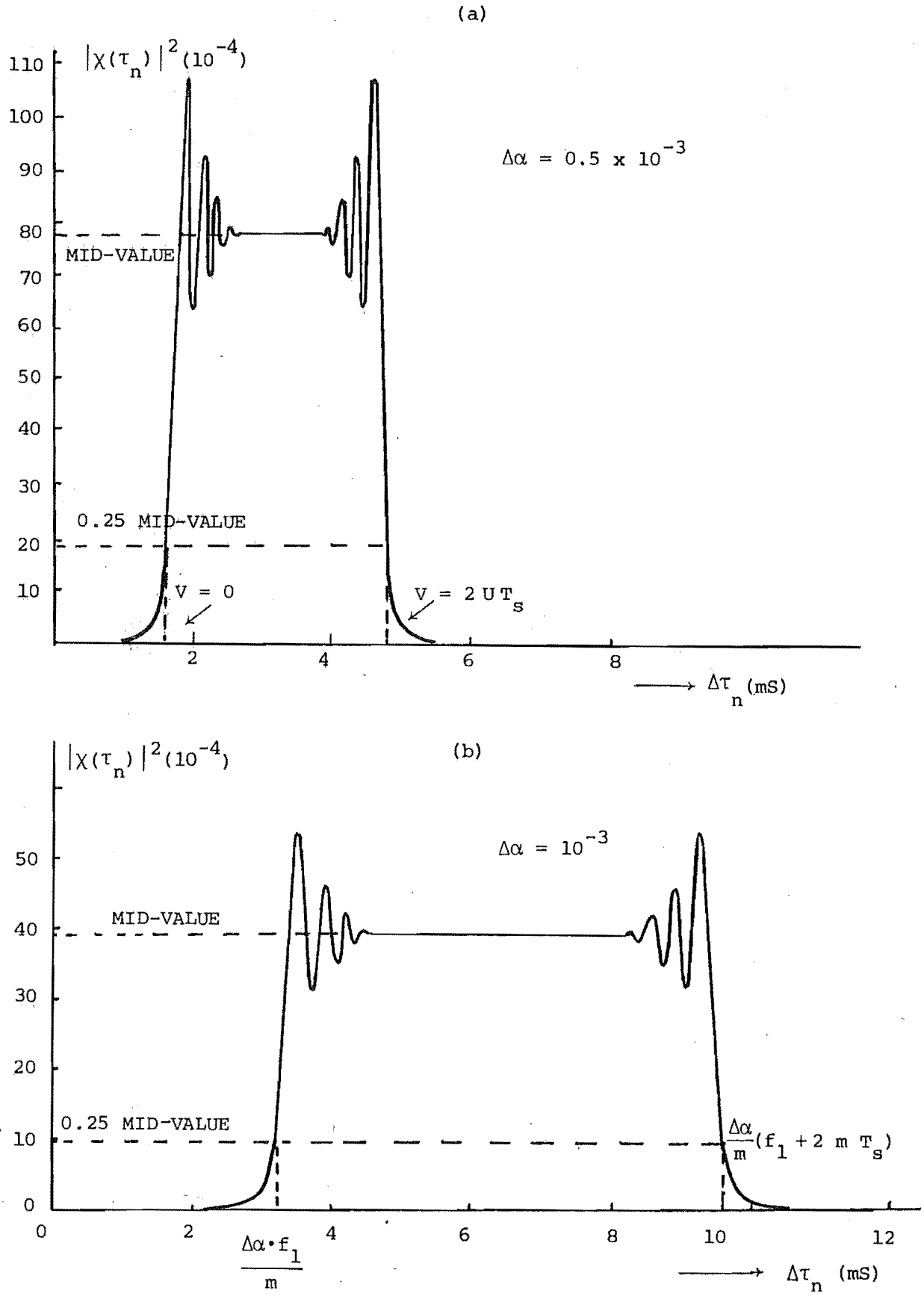
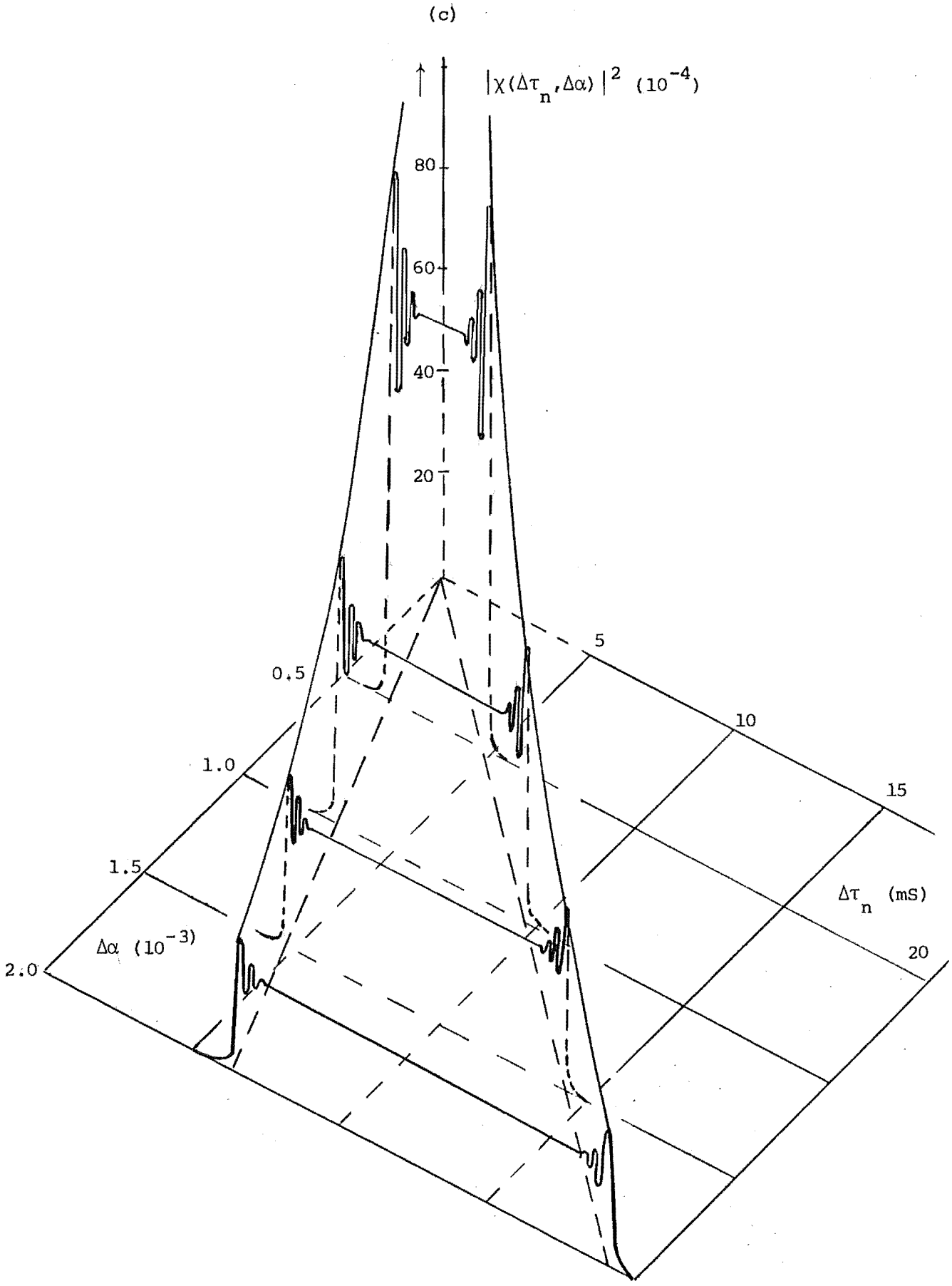


Fig. 3: (a), (b) Cross-sections of the ambiguity function $|\chi(\Delta\tau_n, \Delta\alpha)|^2$ in the constant $\Delta\alpha$ planes.
(c) A three dimensional plot of $|\chi(\Delta\tau_n, \Delta\alpha)|^2$.


(c)
$|\chi(\Delta\tau_n, \Delta\alpha)|^2$ (10^{-4})
80
60
40
20
0.5
1.0
1.5
2.0
$\Delta\alpha$ (10^{-3})
5
10
15
20
$\Delta\tau_n$ (mS)

c) $\Delta\tau_n = 0$, $|\chi(0,\Delta\alpha)|^2$ can be shown to decrease very fast from 1 , at $\Delta\alpha = 0$, to 5.83×10^{-3} at $\Delta\alpha = 0.5 \times 10^{-4}$, and to 1.52×10^{-4} at $\Delta\alpha = 10^{-4}$.

VIII.2 AMBIGUITY CONTOUR AND RESOLUTIONS IN RANGE AND IN VELOCITY

The ambiguity contour of a CTFM wave form is represented by the equation

$$|\chi(\Delta\tau_n,\Delta\alpha)|^2 = \text{Constant } K \tag{24}$$

A two dimensional plot of such an ambiguity contour in the plane $(\Delta\tau_n,\Delta\alpha)$ is called an ambiguity diagram. The resolution in range and in velocity of the system can be determined from this plot. Observing that the ambiguity function, when plotted as a function of $\Delta\tau_n$ for a given value of $\Delta\alpha$, has the gating form defined by its mid value, $(2\ \Delta\alpha\ W\ T_s)^{-1}$, and the limits of

$$\Delta\tau_{n1} = \frac{\Delta\alpha}{m}\ f_1 \tag{25}$$

$$\Delta\tau_{n2} = \frac{\Delta\alpha}{m}\ (f_1 + 2\ m\ T_s) \tag{26}$$

where $|\chi(\Delta\tau_n,\Delta\alpha)|^2$ is equal to 0.25 of the mid value.
Thus the ambiguity contour may be found quickly by setting the mid value equal to K to determine $\Delta\alpha$, then using equations (25) and (26) to determine $\Delta\tau_{n1}$ and $\Delta\tau_{n2}$

$$|\chi(\Delta\tau_n,\Delta\alpha)|^2_{mid} = \text{constant } K \tag{27}$$

$$\Delta\alpha = \frac{1}{2\ K\ W\ T_s} \tag{28}$$

$$\Delta\tau_{n1} = \frac{f_1}{2\ K\ W^2} \tag{29}$$

$$\Delta\tau_{n2} = \frac{1}{2\,K\,W}\left(2 + \frac{f_1}{W}\right) \tag{30}$$

Figure 4 shows an ambiguity diagram plotted by this technique. To determine the resolution in range and in velocity, the constant K is frequently set at 0.25 . This is equivalent to a half value from the complete correlation between two coincident targets (same range and same velocity). Thus the resolution in velocity is determined by

$$\Delta\alpha = \frac{2}{W\,T_s} \tag{31}$$

and the resolution in range is determined by the limits

$$\Delta\tau_{n1} = \frac{2\,f_1}{W^2} \tag{32}$$

$$\Delta\tau_{n2} = \frac{2}{W}\left(2 + \frac{f_1}{W}\right) \tag{33}$$

These equations, (31), (32) and (33), show that the resolution in range (in delay) decreases with increasing the product of the signal bandwidth and the percentage bandwidth, while the resolution in velocity is inversely proportional to the product of the signal bandwidth and its period of repetition. Applying these results in the example of the fishing sonar described in section VIII.1, it is found that $\Delta\alpha = 1.56 \times 10^{-5}$, $\Delta\tau_{n1} = 0.5 \times 10^{-4}$ secs, $\Delta\tau_{n2} = 1.5 \times 10^{-4}$ secs. Typically the sound velocity in water is $C = 5000$ ft/s . Thus the resolution in velocity and in range is determined by: $\Delta v = \frac{C}{2}\Delta\alpha = 4 \times 10^{-2}$ ft/s , $\Delta r_1 = \frac{C}{2}\Delta\tau_{n1} = 0.125$ ft , $\Delta r_2 = \frac{C}{2}\Delta\tau_{n2} = 0.375$ ft.

It is also noted that the above values are not the absolute resolutions of the sonar system. The equations (31), (32) and (33) only provide the limits of the resolution performance of the system in the sense that the cross-correlation between the echoes returned from two nearby targets having that difference in range and in

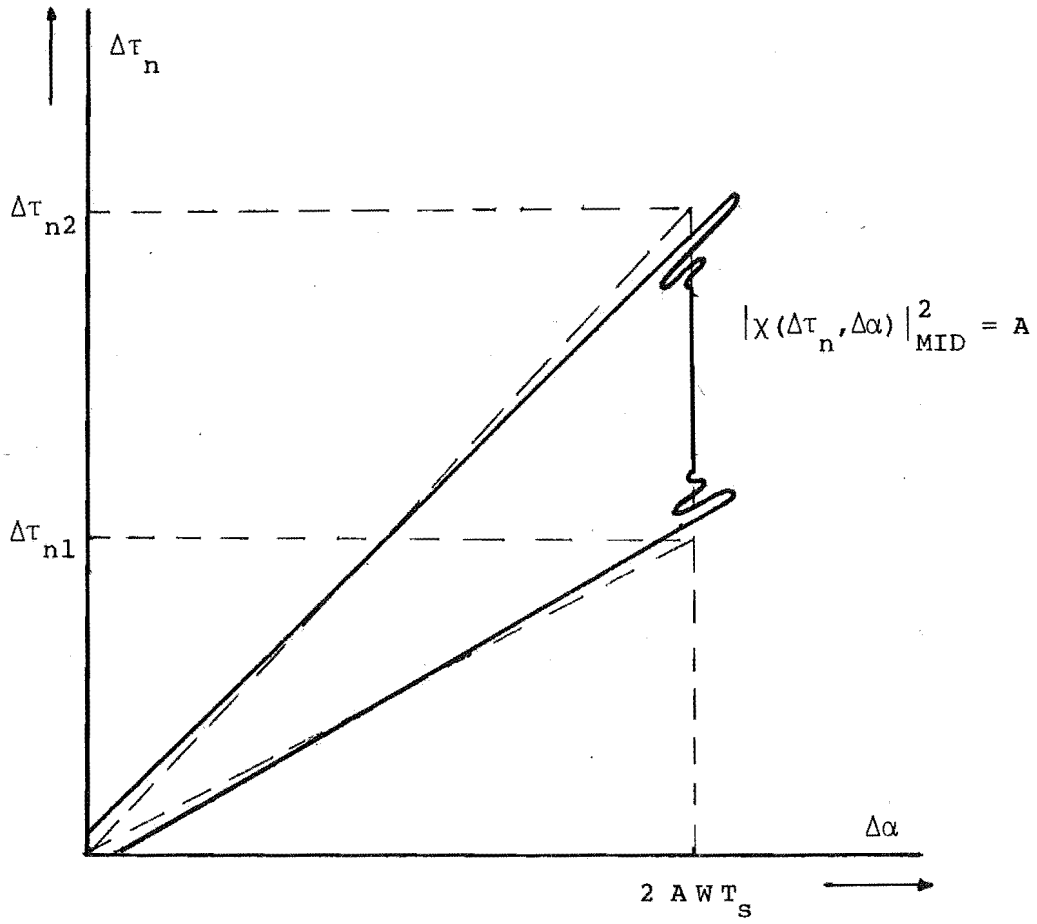


Fig. 4: Ambiguity diagram of wide band CTFM sonar.

velocity is equal to a quarter of the maximum value when two targets are completely coincident

$$\text{i.e.} \qquad |\chi(\Delta\tau_n, \Delta\alpha)|^2 = 0.25$$

$$\text{or} \qquad |\chi(\Delta\tau_n, \Delta\alpha)| = 0.50$$

In practice, the realization of the display system may require consideration. Depending on the specific values to which the maximum target velocities encountered by the system are restricted, large differences in range and in velocity may be required for resolution. That corresponds to a lower value of the cross-correlation coefficient $|\chi(\Delta\tau_n, \Delta\alpha)|^2$.

VIII.3 RELATIONSHIP BETWEEN THE RANGE CODING FREQUENCY AND AMBIGUITY FUNCTION

In chapter 3, it was shown that the range coding frequency of the CTFM sonar system for an object of instantaneous range $r(t)$ and approach velocity v is given by

$$fa(t) = \frac{2m}{C} r(t) - \frac{2v}{C} (f_1 + m\, t_s) \tag{34}$$

where $r(t) = r_o - v\, t$.

Using the symbols defined in section VIII.1 : $\tau_o = \frac{2 r_o}{C}$, $\alpha = \frac{2v}{C}$, ~~$t = t_n + t_s$~~ ~~$t_s = t_n + t_s$~~ , and ~~τ_n~~ $\tau_m = \tau_o - \alpha\, t_n$; equation (34) can be rearranged as follows

$$fa(t) = m\, \tau_n - \alpha\, f_1 - 2\, \alpha\, m\, t_s \tag{35}$$

where $0 \leq t_s \leq T_s$

At the time t , the difference of the range coding frequencies of two nearby objects having the differences in delay and in velocity equal to $\Delta\tau_n$ and $\Delta\alpha$ successively, must be:

$$\Delta f_a(t) = m\,\Delta\tau_n - \Delta\alpha \cdot f_1 - 2\,\Delta\alpha \cdot m\,t_s \qquad (36)$$

$$0 \leq t_s \leq T_s$$

For a certain value of $\Delta\alpha$, the amount of ambiguity in range of the two targets is maximum if $\Delta f_a(t) = 0$, or

$$\Delta\tau_n = \frac{\Delta\alpha}{m}(f_1 + 2\,m\,t_s) \qquad (37)$$

Since at the time t, in the n-th cycle of the sweep frequency, t_s may arbitrarily have any value within the limits from 0 to T_s, therefore $\Delta\tau_n$ falls in the limits

$$\Delta\tau_{n1} = \frac{\Delta\alpha}{m} f_1 \qquad (38)$$

$$\Delta\tau_{n2} = \frac{\Delta\alpha}{m}(f_1 + 2\,m\,T_s) \qquad (39)$$

The equations (38) and (39) are exactly identical to the equations (25) and (26) found previously by means of the ambiguity function. They determine the limits of the difference in range of two objects where the range coding frequencies start to separate from each other.

ACUSTICA
S. HIRZEL VERLAG · STUTTGART
Volume 36 1976/77 No. 1

An Artificially Generated Multiple Object Auditory Space for Use where Vision is Impaired

by L. Kay and M. A. Do

Electrical Engineering Department, University of Canterbury, Christchurch, New Zealand

Summary

A display of spatial information using audible signals with binaural characteristics is described. The spatial information is gathered by a very wide band wide beam Continuous Transmission Frequency Modulated echo-location system, specially designed either for aiding blind persons or for use in fish location. The paper discusses the forms of the echo signals which are produced by different relative motions and shows that the range coding of pitch, being proportional to distance in stationary conditions, changes to a complex sound pattern uniquely related to the motion.

It is suggested that blind people have learned to use these sound patterns, which flow towards the left or the right as objects are passed, rather than the simple range code as described for a static situation. The binaural display resolution is considered under static and dynamic conditions.

Ein künstlich erzeugter Multi-Objekt-Hörereignisraum zur Verwendung bei Behinderung des Sehvermögens

Zusammenfassung

Es wird eine Abbildung räumlicher Information beschrieben, die akustische Signale mit binauraler Charakteristik benützt. Die räumliche Information erhält man durch ein sehr breitbandiges und breitstrahliges FM-Echo-Ortungssystem mit kontinuierlicher Übertragung, das speziell für Blinde und für die Ortung von Fischen entworfen wurde. Der Artikel diskutiert die Formen des Echosignals, die durch verschiedene Relativbewegungen erzeugt werden und zeigt, daß die Bereichscodierung der Tonhöhe, die bei stationären Bedingungen der Entfernung proportional ist, sich zu einem komplexen Klangmuster wandelt, das in eindeutiger Beziehung zur Bewegung steht. Es wird angenommen, daß Blinde den Gebrauch dieser Klangmuster, die nach rechts oder nach links fließen, wenn Objekte sich vorbeibewegen, erlernt haben und nicht die einfache Bereichscodierung, wie sie für eine statische Situation beschrieben wird. Die binaurale Bildauflösung wird unter statischen und dynamischen Bedingungen betrachtet.

Réalisation d'un espace d'écoute artificiel à usages multiples

Sommaire

On décrit une présentation d'information spatiale utilisant des signaux sonores à caractéristiques binaurales. Cette information est recueillie par un dispositif Sonar à faisceau de très large bande, à transmission continue et modulé en fréquence, destiné spécialement soit pour aider des aveugles, soit pour repérer des bancs de poissons. On examine les différentes formes d'échos évoqués suivant les mouvements relatifs, et l'on montre que la bande de codage de la hauteur du son étant proportionelle à la distance lorsque les objets sont stationnaires, on obtient dans le cas du mouvement un son complexe, qui ne dépend que de ce mouvement.

On suggére que les aveugles ont appris à reconnaître ces allures complexes dont les courbes caractéristiques s'écoulent vers la gauche ou vers la droite lorsqu'ils dépassent des obstacles; cette hypothèse paraît plus vraisemblable que celle d'un simple codage en distances, comme on le décrit pour des objets immobiles. On discute du pouvoir de résolution du dispositif, dans les cas statique et dynamique.

1. Introduction

It was first proposed by Kay [1] that an auditory space for use by blind persons could be artificially generated by employing a wide angle ultrasonic radiating field as an "illuminating" source, together with two receivers, one feeding each ear, to convert reflections into audible sounds. These would be perceived binaurally. The distance to a reflection was to be coded in the form of rising pitch with increasing distance (frequency proportional to distance), and the binaural differences, which could include time, frequency and amplitude, would indicate the direction of a reflection. Subsequently, Kay, [2] found that amplitude difference was the dominant cue. It was later demonstrated by Rowell [3] that time and frequency differences were in-

compatible, impeding the fusion of binaural signals, and proposed that the possible interaural differences be reduced to that of only amplitude — as far as this was physically possible.

The means for producing this new form of auditory space was built in the form of a sensory aid for the blind, shown in Fig. 1 [4], [5], and a sonar to locate fish, Fig. 2 [6], [7].

Both systems were evaluated during 1970—72.

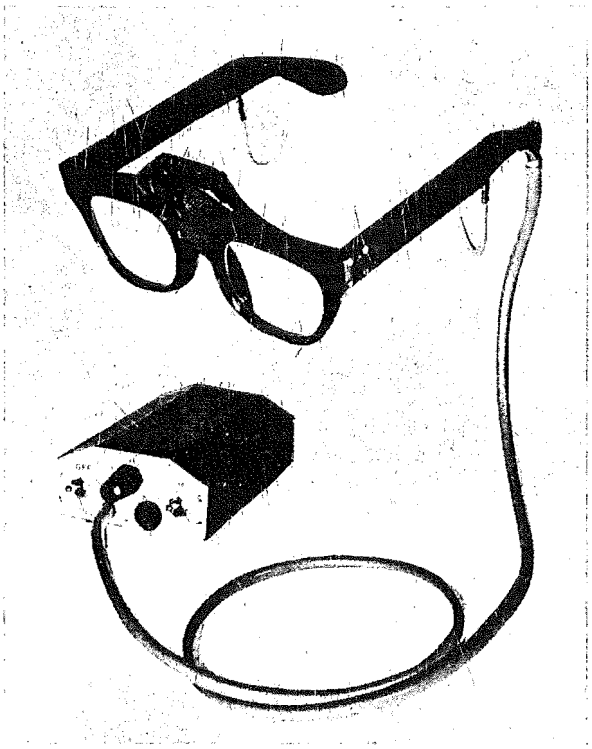


Fig. 1. Binaural sensory aid for the blind.

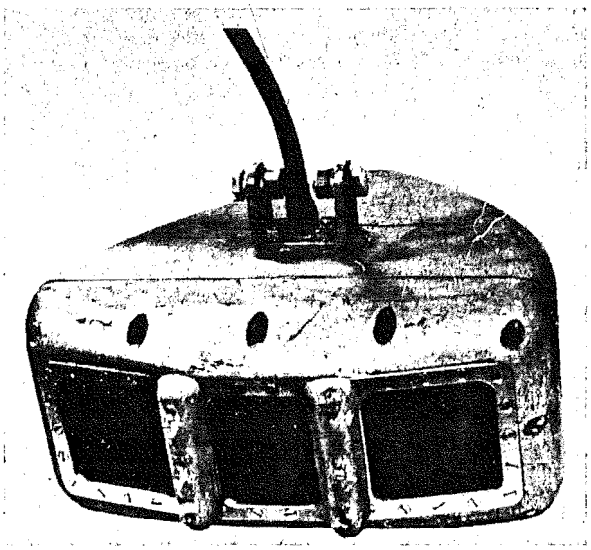


Fig. 2. Transducer for binaural sonar.

2. Evaluation

2.1. Sensory aid for the blind

About 200 blind people have learned to use the sensory system as an aid to spatial perception and to gain improved mobility. When 100 people had been trained by specially qualified teachers, both the users and teachers were questioned by a professionally prepared questionnaire [8], [9]. It was found that users learned to interpret the very rich auditory information with comparative ease. After 40 to 60 hours of training over a period of four weeks, many of the users were able to travel in a busy pedestrian area with grace and confidence not seen in totally blind people before. A few, at least, approached a behaviour pattern not unlike that of a sighted pedestrian (as recorded on film).

Two schools[1], specialising in the training of mobility instructors and who were involved in the evaluation, have now commenced training more instructors in the use of the Binaural Sensory Aid. This form of auditory space has come to be accepted as a viable means for aiding blind people to perceive their immediate environment.

3.2. Fishing sonar

The fishing sonar was fitted to the bows of a 5.4 m boat and tested in 36···54 m of water in the Marlborough Sounds (N.Z.). Even when the water was choppy with 0.3 m waves causing the boat to rock and pitch considerably, it was easy to track a 45 cm dimater sphere submerged to 18 m as the boat and sphere drifted apart up to 100 m. The transducers was not stabilised; its field of view covered an arc of 50° in azimuth and 30° in evaluation. The listener was able to correlate the movement of the boat with the changes in auditory sounds and "stabilise" his auditory space. This was the most unexpected of the findings using the new display. Earlier trials with a stabilised F. M. Sonar had demonstrated the capability to the auditory system to track similar underwater targets [10].

Later sea trials lasting four days on a 14 m launch, again with the transducer fitted to the bows, showed that shoals of fish could be readily detected up to 135 m. The boat was easily steered by an untrained person over the top of the shoal so as to obtain confirmatory evidence on an echo-sounder. When two shoals were located at the same time in different

[1] Dept. of Blind Rehabilitation, College of Education, Western Michigan University, Kalamazoo, Michigan 49001, U.S.A., and National Guide Dog Training Centre, Royal Guide Dogs for the Blind Associations of Australia, P.O. Box 162, Kew, Victoria 3101, Australia.

directions and at different distances, the choice could be made to steer over either the larger or the near shoal. Greater ranges are thought likely if the transducer were fitted deeper under the hull of the boat and some engineering modifications made to the transducer design.

It was also found that two different species of fish could be recognised as "different" by the difference in the character of the echo sounds they produced. The probability of locating a shoal appeared to be increased significantly [7].

2.3. *Discussion*

The results of both the evaluations indicated clearly the ability of human operators to use the new auditory display with relative ease, even though the input to the ears was rich in information. Whilst the sonar operator steering a boat travelling at 2.5 m/s has approximately 60 s in which to carry out his operation, a blind person walking at about 5 km/h along a footpath has only 2.0 s in which to avoid a stationary pedestrian in his path at 3 m distance. The blind traveller then experiences exceptionally high rates of fractional range change. Even so, the control functions are readily generated by the user so that the task is easily executed [11].

Under these dynamic conditions the display is not correctly described by the very simple concept presented in the introduction. This paper describes the auditory signals more completely and explains why simple controlled laboratory experiments may give misleading results.

3. System concept

The basic parameters of the auditory display of spatial information are simply described under stationary conditions — the distance to an object in the system field of view, of say 60°, is proportional to the frequency of the audible sound output of the device (typically 900 Hz/m for the Blind Aid and 25 Hz/m for the sonar), and the direction of an object is indicated by the binaural difference in loudness of the sounds fed to each ear (typical I.A.D. is 0.4 dB per degree in both systems). These are illustrated in Fig. 3.

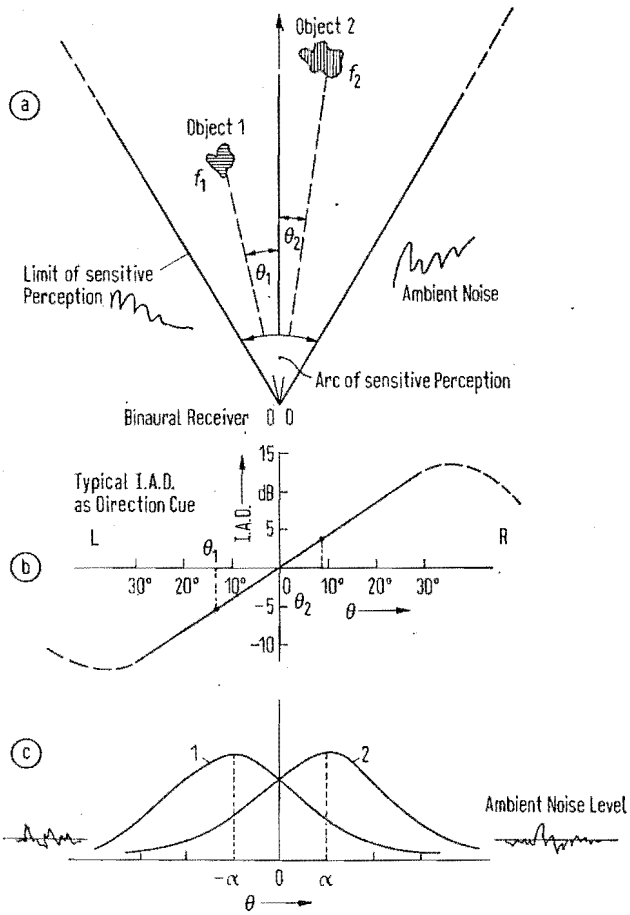


Fig. 3. Illustration of basic parameter for sonar. Maximum range is 180 m for fish sonar, and 6 m for blind aid, the corresponding maximum audio frequency is 5 kHz.
(a) Targets in sonic beam,
(b) binaural direction code, I.A.D. proportional to θ,
(c) the receiver responses: Left curve (1) and right curve (2).

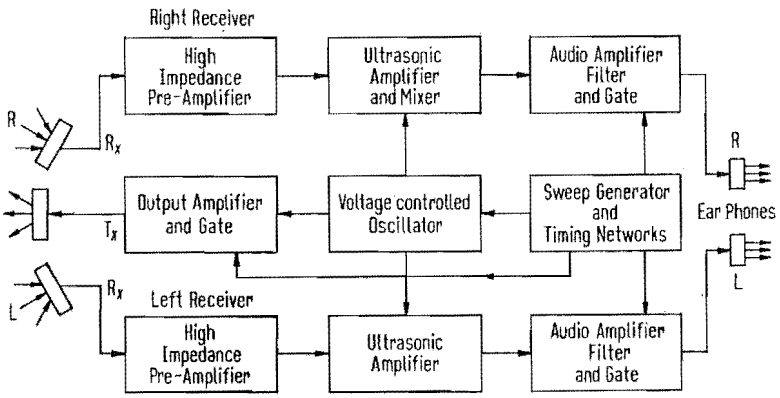


Fig. 4. Block diagram of a sonar system.

To provide these combined distance and direction indicators, a continious transmission frequency modulated echo location system must be used, typically — but not restricted to — operating over a frequency range of approximately 40 kHz to 80 kHz as shown in Fig. 4. There seems to be no other simple way to obtain these display parameters which are quite unique in their more complex form under the dynamic conditions of human location.

The radiated signal of the transmitter is cyclic and of the form (shown in Fig. 5)

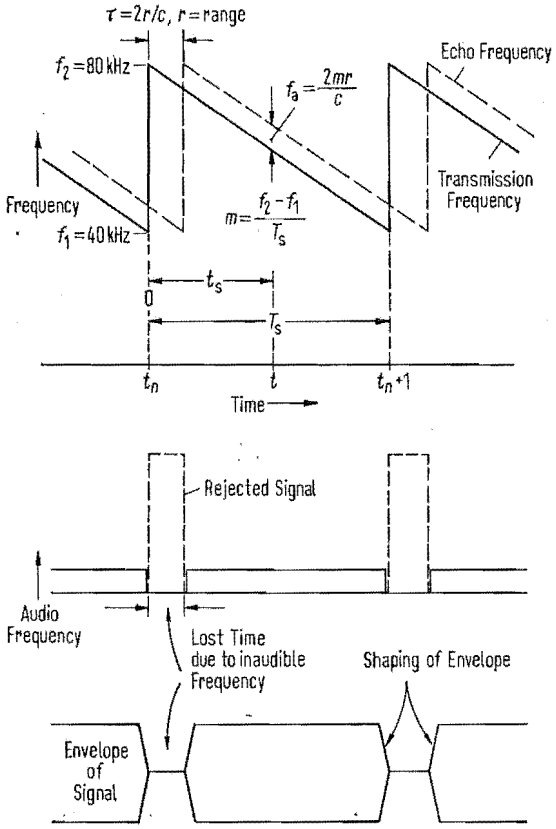


Fig. 5. F.M. sweep parameters and audio output signal (stationary condition).

$$S_T(t) = A \exp 2\pi j\,(f_2 t_s - \frac{m}{2} t_s^2) \qquad (1)$$

where

t_s is the time during a sweep cycle, i.e. $0 \leqq t_s \leqq T_s$,
T_s is the sweep period,
m a constant to be chosen.

This signal may be assumed to have a rectangular envelope of duration T_s neglecting the "lost time", Any instant, $t = t_n + t_s$; where t_n is the beginning of the n-th period of the modulating function and t_s varies between 0 and T_s.

At the time t, the instantaneous transmitted frequency is

$$f_T(t) = f_2 - m t_s . \qquad (2)$$

3.1. *Stationary conditions*

Under perfectly stationary conditions the received signal $S_R(t)$ is a delayed replica of the transmitted signal reflected by an object at distance r. Hence

$$S_R(t) = \alpha(r)\, S_T(t - \tau) . \qquad (3)$$

If c is the velocity of sound in the propagating medium then τ, the range delay, is $2r/c$; $\alpha(r)$ is the attenuation due to range r. The instantaneous phase angle of the received signal is then

$$\theta_R(t) = 2\pi (f_2 (t_s - \tau) - m/2 \cdot (t_s - \tau)^2) . \qquad (4)$$

The received frequency is given by

$$\frac{1}{2\pi} \frac{d\theta_R(t)}{dt} = f_R(t) = f_2 - m\left(t_s - \frac{2r}{c}\right). \qquad (5)$$

The range code of the system is obtained from the difference between the transmitted and received frequencies at time t.

$$f_a = f_R(t) - f_T(t) = \\ = \frac{2mr}{c} \qquad (6)$$

and is made audible by a suitable choice of m.

Thus, whilst the transmitted and received signals are time varying and inaudible between 40 and 80 kHz, the audible signal is constant, proportional to the distance r, in the interval $t_n < t < t_{n+1}$ (neglecting short transient intervals $\ll T_s$).

If an object consists of a number of surfaces each scattering energy back to the receiver and the distance to each surface is $r + \Delta r_i$ the audible signal corresponding to the object will be

$$S_a = \sum_{i=1}^{n} \alpha(r)\, A_i \exp\left\{2\pi j \frac{2m}{c}(r + \Delta r_i)\right\} t . \qquad (7)$$

This is a narrow band of tones representing each reflecting point on the surface forming a spectrum of signals uniquely related to the geometry of the object in space, and consequently having a unique audible character.

3.2. *Non-stationary conditions*

3.2.1. Constant velocity situation

When a sensory system is in use either worn on the head or fitted to a boat, there is normally continuously varying relative movement between the environment and the transducers producing a

Doppler shift in the reflected signals. The simple range code described for perfectly stationary conditions then changes to a pattern of sound variation uniquely related to the relative change in distance and direction taking place in the environment.

If, to simplify the situation, the relative velocity between user and an object in space is assumed to be constant (v), then the distance to the object is given by

$$r(t) = r(0) - vt$$

(assuming a radial approach velocity)

where $r(0)$ is the initial range at $t_0 = 0$.

The received signal is then a Doppler shifted "replica" of the transmitted signal τ seconds ago which was reflected from an object at distance

$$r_{(t-\tau/2)} = r_{(t)} + \frac{v\tau(t)}{2} \quad \text{(see Fig. 6)}.$$

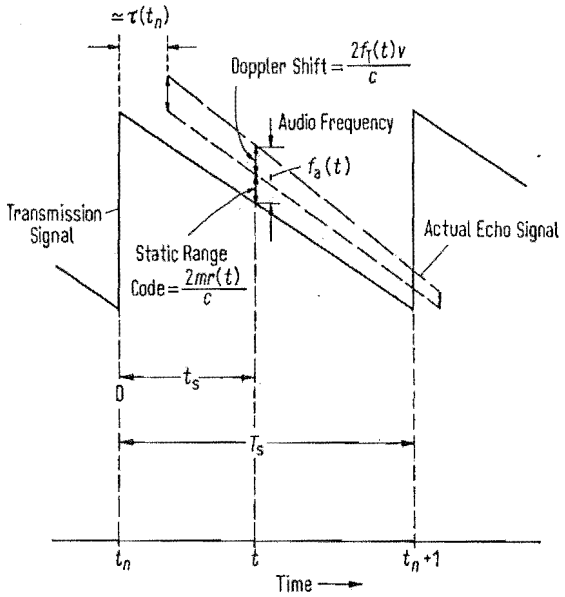


Fig. 6. Doppler shifted "replica of transmission wave" (constant approach velocity $r(t) = -v$).

Then

$$\frac{\tau(t)}{2} = \frac{r + \frac{v\tau(t)}{2}}{c}$$

or

$$\tau(t) = \frac{2r(t)}{c - v} = \frac{2(r(0) - vt)}{c - v}. \tag{9}$$

The received signal is

$$S_R(t) = \alpha(r)\, S_T(t - \tau(t)). \tag{10}$$

Substitution of eq. (1) in eq. (10) leads to the difference frequency between this received signal and the transmitted signal at the same instant and may be shown, by simple differentiation and manipulation, to be

$$f_a(t) \simeq \frac{2m}{c}(r(0) - vt) + \frac{2v}{c}(f_2 - m t_s). \tag{11}$$

To appreciate the significance of this, an example is shown plotted in Fig. 7. There is a shift in the indicated range about which there is a saw-tooth cyclic variation. This pattern of change is due both to the Doppler shift of frequency in the medium and the wide variation in transmitted frequency.

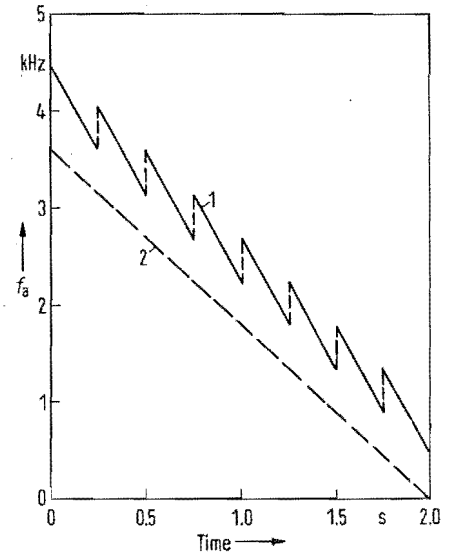


Fig. 7. Audio signal variation due to constant radial velocity v; $v = 1.5$ m/s, $f_2 = 100$ kHz, $f_1 = 50$ kHz, $m = 200$ kHz/s, $T = 250$ ms. The dashed line corresponds to the static range code (i.e. no Doppler effect).

3.2.2. Varying velocity situation

A simple constant velocity situation is rare under conditions of human locomotion — the human body is capable of considerable acceleration, and varying velocity could be considered to be the more normal condition generated by man's peripatetic movement.

It is shown in Appendix I that the general form for the audible signal frequency is

$$f_a(t) \simeq \frac{2m}{c} r(t) - \frac{2 f_T(t)\, \dot{r}\ (t)}{c}. \tag{12}$$

For the special case of constant acceleration, a towards the observer the audible signal frequency

$$f_a(t) \simeq \frac{2m}{c}\left(r(0) - v_0 t - \frac{a}{2} t\right) + \\ + \frac{2}{c}(f_2 - m t_s)(v_0 + at) \tag{13}$$

where v_0 is the initial velocity.

From the plot of Fig. 8 it will be seen that the acceleration produces a non-linear change of audio frequency during the period $t_n < t < t_{n+1}$.

In the case of a fishing sonar the effect of ship's acceleration is usually trivial, but Doppler shift produced by a large fish may not be.

It will be evident that under real life conditions of, for example, mobility by the Blind in a busy pedestrian area where people are moving with varying acceleration, and where the relative velocity between user and fixed objects will also vary, the audio signals from this sensory system will change in an indescribably complex way — each signal having its own unique character (see Fig. 9).

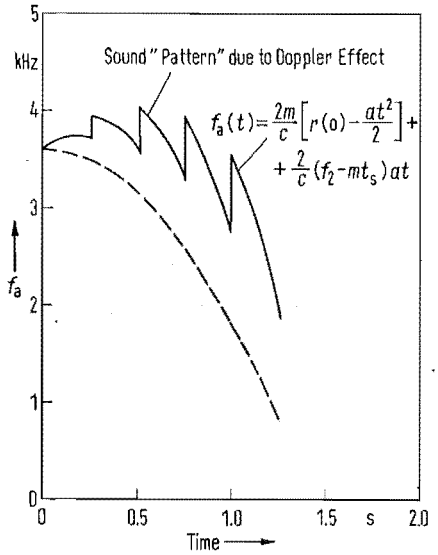


Fig. 8. Constant acceleration sound pattern; Initial velocity $v_0 = 0$; constant acceleration $a = 3$ m/s², $f_2 = 100$ kHz, $f_1 = 50$ kHz, $m = 200$ kHz/s, $T = 250$ ms. The dashed line is the static range code (i.e. no Doppler effect).

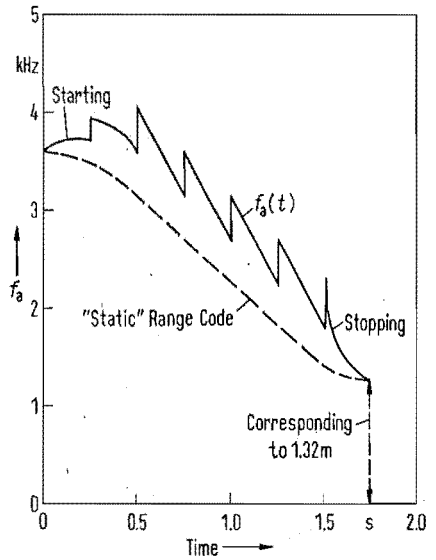


Fig. 9. Varying velocity sound pattern, approaching an object from 3 m and stopping at 1.32 m. For $0 \leq t \leq 0.5$ s, initial velocity = 0, acceleration = 3 m/s²; for 0.5 s $\leq t \leq$ 1.5 s, constant velocity = 1.5 m/s; for 1.5 s $\leq t \leq$ 1.75 s deceleration = 6 m/s².

4. Direction code

The discussion has thus far been about the distance code. The direction code of the binaural sensory system is designed to follow the relationship of

$$\theta_e = k \log (I_R / I_L)$$

as far as this is physically possible — shown plotted in Fig. 3, where

θ_e is the estimated direction as perceived by the user,

k is the auditory localizing coefficient of the user

I_R, I_L the sound intensities of the right and left ears respectively.

Subjects vary in their sensitivity to interaural amplitude difference, and by a suitable choice of k the direction code can be matched to the user so that $\theta_e = \theta$ the actual direction.

The I.A.D. code is obtained in practice by using a specially designed transmitter-receiver transducer arrangement (Kay [5]). The ideal overall transducer response is shown by Rowell [3] to be

$$A(\theta) = c \exp[(-\theta \pm \alpha)^2 / 4\alpha k],$$

where α is the splay angle of the receiver transducers, c a constant (illustrated in Fig. 3). This in practice may only be approximated.

When the two signals of nearly the same frequency are being received, one from object (a) and the other from object (b) separated by an angle Φ the auditory system is unable to discriminate the L and R stimulus from object (a) and the L and R stimulus from object (b). Only when the L and R responses from object (a) fuse to indicate an object in direction $\Phi_{(a)}$ and the L and R responses from object (b) fuse to indicate an object in direction $\Phi_{(b)}$, is it possible to expect discrimination in direction. It is also well known that two tones presented to a subject simultaneously must differ considerably in frequency to be heard as two distinctly separate notes.

It would seem therefore that spatial resolution in the binaural system would be exceedingly poor.

5. Discussion on resolution

It has been shown however that locomotion, when wearing the sensory aid, produces rapidly varying auditory signals — not tones. Only rarely are we interested in the system resolution when standing

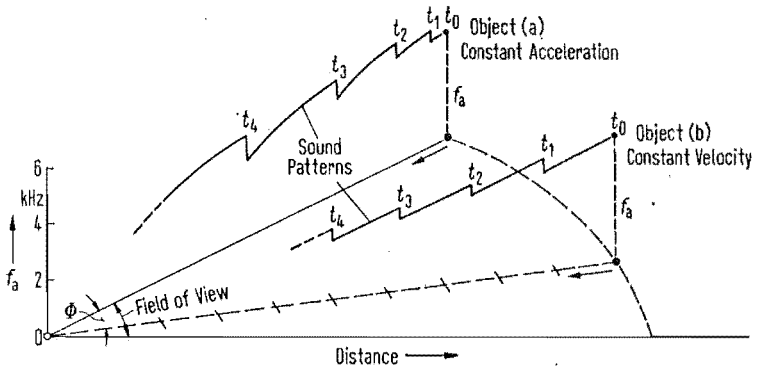


Fig. 10. Perspective of sound pattern for radial motion.

perfectly still. The same applies to the fishing sonar —if less so. Early experiments to determine the resolution as a function of the range and direction cues combined showed that the frequency of two tones presented simultaneously had to be almost an octave apart before they could be heard as distinctly separate and a direction attributed to them [3]. Experience with blind people using the sensory aid in real life situations suggests that much greater resolution is being enjoyed. What blind people must perceive are time varying "tones" producing flow patterns of sound related to the changing spatial positions of the reflecting objects relative to the user. A very simple illustration is given in Fig. 10 using the signals illustrated in Figs. 7 and 8.

The flow patterns obtain when, for example, objects are being approached and then passed on the left or right are not unlike the flow patterns experienced by a driver of an automobile passing street lights in a thick fog. Gibson [12] discusses such flow patterns in vision during locomotion. Thus even though the form of the "sound patterns" may be mathematically very complicated for realistic mobility situations when several objects can be in the field of view at once, each pattern must be perceived as having a very simple cognitive form related to the changing spatial co-ordinates.

It is now believed that these sound patterns each having unique character in both flow and pitch quality (see eq. (7)), such as illustrated in Fig. 10, allows the auditory system of humans to exhibit greater resolution than has previously been shown possible from experiments; it could be the explanation why some blind people have come to use the sensory system so effectively.

To describe quantitatively the resolution capability of the sensory system as a whole, a series of controlled experimental tests have been conducted simulating dynamic conditions. These are reported in the accompanying paper.

(Received August 5th, 1975.)

Appendix I

At the time t, the instantaneous phase angle of the L.F.M.-C.W. transmitted signal,

$$S_T(t) = \exp[|j\,\theta_T(t)|],$$

is represented by

$$\theta_T(t) = 2\pi \left| f_2 t_s - \frac{m}{2} t_s^2 \right| \qquad (A1)$$

where $t = t_n + t_s$, t_n is the beginning of the n-th cycle of period T_s. Thus, while t varies between t_n and t_{n+1}, t_s varies from 0 to T_s; m is the sweep rate, and the instantaneous transmitted frequency is

$$f_T(t) = f_2 - m t_s \quad 0 \leqq t_s \leqq T_s\,, \qquad (A2)$$

f_2 is the upper limit of the transmitted frequency.

If the received signal is a Doppler shifted replica of the transmitted signal with a time delay $\tau(t)$, then its phase angle at the time t is

$$\theta_R(t) = 2\pi \left| f_2(t_s - \tau(t)) - \right.$$
$$\left. - m/2\,(t_s - \tau(t))^2 \right|. \qquad (A3)$$

The received frequency is given by

$$f_R(t) = f_2 - f_2\,\dot{\tau}(t) - m(1 - \dot{\tau}(t))\,(t_s - \tau(t)) \qquad (A4)$$

and the range coding frequency is the difference between $f_R(t)$ and $f_T(t)$

$$f_a(t) = m\,\tau(t)\,(1 - \dot{\tau}(t)) - \dot{\tau}(t)\,(f_2 - m t_s) \qquad (A5)$$

where the dot represents the time derivative.

The general expression of time delay is

$$\frac{\tau(t)}{2} = \frac{1}{c}\, r\left(t - \frac{\tau(t)}{2}\right) \qquad (A6)$$

where $r(t)$ is the instantaneous range, and c is the sound velocity. If the rate of change of $r(t)$ is much smaller than c, then the following approximation is available,

$$\tau(t) \simeq \frac{2\,r(t)}{c}, \qquad \left|\frac{2\,\dot{r}(t)}{c}\right| \ll 1\,,$$

and

$$f_a(t) \simeq \frac{2m}{c} r(t) - f_T(t) \frac{2\dot{r}(t)}{c}. \qquad (A7)$$

References

[1] Kay, L., A new or improved apparatus for furnishing information as to position of objects. Patent Specification No. 978741. The Patent Office, London 1959.

[2] Kay, L., Blind Aid. Patent Specification No. 3366922. United States Patent Office, Washington, D.C. 1965.

[3] Rowell, D., Auditory display of spatial information. Ph. D. Thesis, University of Canterbury 1970.

[4] Martin, G., Electronics and transducers for an ultrasonic blind mobility aid. M.E. Thesis, University of Canterbury 1969.

[5] Kay, L., A sonar aid to enhance spatial perception of the blind: Engineering design and evaluation. Radio Electron. Eng. 44 [1974], 605.

[6] Smith, R. P. and Kay, L., A fishfinding sonar utilizing an audio information display. Digest of Technical Papers, IEEE Ocean Conf., Panama City, Florida 1970, p. 113.

[7] Smith, R. P., Transduction and audible displays for broad band sonar systems. Ph. D. Thesis, University of Canterbury 1973.

[8] Airasian, P., Evaluation of the binaural sensory aid for the blind. A.F.B. Res. Bull. 26 [1973], 51.

[9] Kay, L., Sonic glasses for the blind — Presentation of evaluation data. A.F.B. Res. Bull. 26 [1973], 35.

[10] Kay, L., An experimental comparison between pulse and frequency-modulation echo-ranging systems. J. Brit. I.R.E. 19 [1960], 785.

[11] Kay, L., Toward objective mobility evaluation, some thoughts on a theory. A.F.B. New York, monograph 1974.

[12] Gibson, J. J., Visually controlled locomotion and visual orientation in animals. Brit. J. Psychol. 49 [1958], 182.

ACUSTICA
S. HIRZEL VERLAG · STUTTGART
Volume 36 1976/77 No. 1

Resolution in an Artificially Generated Multiple Object Auditory Sensations Space Using New Auditory Sensations

by M. A. Do and L. Kay

Electrical Engineering Dept., University of Canterbury, Christchurch, New Zealand

Summary

Auditory space, produced by means of a Continuous Transmission Frequency Modulated (C.T.F.M.) wide beam, wide band sonar, has been used in a sensory aid for the blind for some time. Yet, no satisfactory definition of auditory resolution which is directly related to spatial resolution, has been obtained and measured by controlled experiments. In this paper, laboratory experiments, using realistic system simulation techniques are described which determine the ability of subjects to resolve objects in space by auditory means, under both static and dynamic conditions. It is shown that while the auditory frequency resolution is poor under static conditions, it is significantly improved in the dynamic case. The Doppler effect, involving doubling the rate of change of the audio frequency, plays a significant part in this improvement. An increase in relative velocity of an object may also enhance the resolution capability quantitatively as well as qualitatively to an extent that signals may be resolved cognitively before the subject's reaction can take place.

Auflösung in einem künstlich erzeugten Multi-Objekt-Hörereignisraum unter Verwendung neuer Hörempfindungen

Zusammenfassung

Ein durch ein breitbandiges und breitstrahliges FM-Sonar-System mit kontinuierlicher Übertragung (C.T.F.M.) erzeugter Hörereignisraum wurde seit einiger Zeit in einer Sehhilfe für Blinde benutzt. Bis jetzt wurde noch keine befriedigende Definition der Hör-Auflösung, die in direkter Beziehung zur räumlichen Auflösung steht, angegeben und in kontrollierten Experimenten gemessen. In dieser Arbeit werden Laborexperimente beschrieben, die unter Verwendung realistischer System-Simulationstechniken die Fähigkeit von Versuchspersonen, Objekte im Raum über den Gehörsinn zu lokalisieren und zu unterscheiden, sowohl unter statischen als auch unter dynamischen Bedingungen bestimmen. Es wird gezeigt, daß die gehörmäßige Frequenzauflösung unter statischen Bedingungen gering ist, während sie sich im dynamischen Fall signifikant verbessert. Der Doppler-Effekt, der eine Verdopplung der Hörfrequenzänderung mit sich bringt, spielt bei dieser Verbesserung eine wesentliche Rolle. Ein Anstieg der Relativgeschwindigkeit eines Objekts kann auch die Auflösungsfähigkeit sowohl quantitativ als auch qualitativ so erhöhen, daß Signale kognitiv aufgelöst werden können, bevor die Reaktion der Person stattfindet.

La résolution dans un espace d'écoute artificiel à usages multiples utilisant des sensations auditives nouvelles

Sommaire

On a utilisé un espace d'écoute produit par un large faisceau de transmission modulé en fréquence formant un système sonar (CTFM), depuis quelque temps déjà, comme aide aux aveugles; mais on n'a pas défini de façon satisfaisante le pouvoir de résolution auditif lié directement au pouvoir de résolution spatial, et on ne l'a pas soumis a des mesures de contrôle. Dans le présent article, on décrit des expériences de laboratoire, utilisant des techniques réalistes de simulation; ces expériences déterminent la capacité des sujets à distinguer des objets dans l'espace par des moyens auditifs, dans des conditions tant statiques que dynamiques. On montre que si la résolution en fréquences audible est médiocre dans des conditions statiques, elle est notablement meilleure dans le cas dynamique. L'effet Doppler, qui double le gradient de variation de la fréquence sonore, joue un rôle important dans cette amélioration. Le pouvoir de résolution peut aussi être amélioré, quantitativement et qualitativement, par un accroissement de la vitesse relative de l'obstacle, à tel point que les signaux peuvent être résolus de manière cognitive avant que la réaction du sujet n'ait pu se produire.

1. Introduction

In the preceding paper by Kay and Do [1], it was shown that auditory signals produced by a C.T.F.M. wide-beam, very wide-band sonar vary in a manner which is uniquely related to spatial change. It was suggested that the patterns of change which are produced by relative movement could be the reason why blind users of this sensory system are able to discriminate the complex signals from multiple objects in a real environment. This was deduc-

ed from the observed ability of blind people to negotiate complex real environments and be more effectively mobile [2]. No supporting measures of a quantitative nature were obtained during the extensive evaluation [3] because of the difficulty in collecting meaningful data on what subjects were using as spatial cues under dynamic conditions. It is of course possible to measure performance in a simple well controlled situation, [4] but once an individual is allowed free movement and personal control of his motion in his normal habitat, the variables become impossible to handle.

Laboratory experiments prior to the evaluation, which were designed, using system simulation, to determine the ability of subjects to resolve objects in space by auditory means, used only static situations to facilitate controlled measurement [5]. The results of these experiments showed subjects had poor resolution capability and failed to explain subjective impressions of resolution capability whilst in motion when using the sensory system. Rowell [5] did, however, show that the simulated direction cue of interaural amplitude difference could be matched to individuals more effectively when rotational head movement was permitted to vary the cue. Subjects could estimate direction more accurately when relative angular motion took place. Time did not permit other forms of motion to be simulated.

In the case of the fishing sonar designed by Smith [6] the ability of an operator to discriminate between two shoals of fish appearing simultaneously in the field of view remains unexplained. Here the operator is presented with reverberation from all ranges simultaneously, the "cluster" of tones — one from each fish — in shoal (a) in the mean direction θ_a and the "cluster" of tones from shoal (b) in the mean direction θ_b. He was able to say which shoal was the larger and which was the nearer of the two, Experiments to determine auditory resolution under these conditions will require the generation of many tones having slightly differing spatial characteristics which include the random spatial motion of fish within the shoal. Production of such a set of conditions has not yet been attempted.

This paper describes some simple experiments using simulation techniques as a first attempt to determine the apparent influence of spatial change on resolution in an artificially generated auditory space. Only two objects are simulated and the relationship to the sensory aid for the blind and the fishing sonar are discussed. A new auditory sensation is produced.

2. Frequency resolution of the ear – the static situation

Studies in auditory perception using two tone are generally referred to as masking experiment rather than resolution experiments because of th nature of the enquiry. Here we are interested in hov well we can perceive space by auditory means an determine how far apart objects must be in order t perceive two separate things. This concept has bee extensively studied in relation to visual displays c space such as used in pulsed sonars and radars, an physical criteria for design purposes have been de veloped. No similar design criteria exists in auditor perception because no satisfactory physical mode exists for the auditory process.

Plomp and Steeneken used two response criteri for their masking experiments which have littl physical meaning but which seem to be readil perceived by the subjests as psychophysical sensa tions. These were a) the frequency difference be tween two tones for which the harshness of dis sonance produced by the tones reaches a maxi mum, and b) the frequency difference for jus absence of interference between two tones Neither of these criteria lead to an understanding c auditory resolution as we seek to define it. Rowe [5] sought to define the frequency resolution as th least frequency difference between two tones s that one could perceive both clearly and distinctly Using six subjects he showed that judgments coul vary from $\Delta f/f$ being 4% to 40% where Δf is th frequency difference and f the mean of the two fre quencies. When questioned, those subjects whic gave 40% said they heard the two tones as dis tinctly separate. This latter result was th closest to indicating the resolution capability of th auditory system in terms which related to the sona system we were attempting to evaluate.

A procedure has now been designed to measur the frequency resolution which is defined as th least frequency difference between two tones so tha they are perceived as being distinctly separate

2. 1. Procedure

Let f_1 be the primary frequency, and f_2 ($< f_1$) b the secondary frequency. Initially $f_2 \ll f_1$ so tha they are distinctly heard as two separate tones. f_2 i then increased step by step until it reaches th maximum value for which both f_1 and f_2 are sti resolvable in the sense defined. For each pair of f and f_2, two tests are successively carried out t verify the subjects' judgment that

1) Two and only two tones are perceived,

2) the perceived tones are the stimuli.

Test 1. In this test the subject's task is to listen to a sound, which randomly includes either two components f_1 and f_2 only, or three components f_1, f_2 and f_3, where the additional component f_3 may have one of the values $2f_2 - f_1$, or $f_1 - f_2$. A "Yes" or "No" answer must be given to the question "is this only *two tones*". A "No" answer should be given when three components are included in the stimuli.

Test 2. If the answers given to a pair of f_1 and f_2 are correct, test 2 is carried out. Right after a "Yes" answer is given in test 1, a single tone is then used as a stimulus through the subject's headphones. Either of two questions are then asked "is this the low tone", or "is this the high tone", while this single tone may have the value of either f_1, f_2, $2f_2 - f_1$, or $f_1 - f_2$. If the answers are correct, the tests are repeated with another pair of f_1 and f_2.

The experiment was carried out in an anechoic chamber and equipment was set up as in Fig. 1.

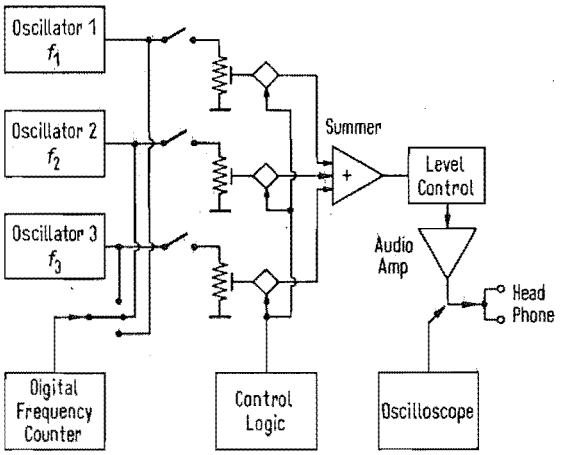


Fig. 1. Equipment arrangement for resolution in the static situation (Experiment 1).

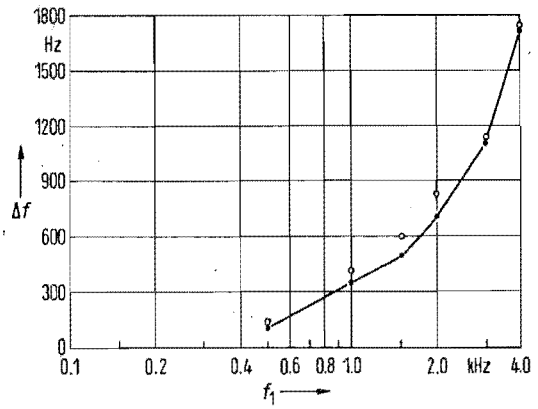


Fig. 2. Mean frequency difference so that two tones can be resolved.
●——● Continuous tone,
○ ○ ○ pulsed tone (period 2s, blanking 0.2s).

The mean values given by four subjects are presented in Fig. 2. Both continuous and pulsed tones were used. Taking $f = (f_1 + f_2)/2$ and $\Delta f = f_1 - f_2$, then $\Delta f/f$ was approximately 40%. These were in agreement with the findings of Rowell.

3. Resolving two changing tones – the dynamic situation

Consider now the simulation of two adjacent objects, 1 and 2, in auditory space described by two pairs of variables $[r_1(t), \theta_1(t); r_2(t), \theta_2(t)]$ where $r(t)$ and $\theta(t)$ are the instantaneous object range and azimuth angle respectively. Motion of an object with respect to the sonar system (Blind Aid or Fish Sonar) changes the frequency of the transmitted wave as it is reflected by a factor of approximately $(1 - 2\dot{r}(t)/c)$ where $\dot{r}(t)$ is the relative velocity of the object and c is the velocity of propagation of the sound wave. Then for a broadband transmission signal higher frequencies must undergo larger Doppler shifts than the low frequencies in the transmission band. The range coding of the sonar system for an object of range $r(t)$ is given by (see Kay and Do [1], appendix 1).

$$f_a(t) \cong \frac{2m\,r(t)}{c} - 2f_T(t)\frac{\dot{r}(t)}{c} \qquad (1)$$

where

$\dot{r}(t)$ is the time-derivative of $r(t)$,
m is the sweep rate of the modulated transmission frequency $f_T(t)$, $f_T(t) = f_2 - mt_s$;
f_2 is the upper limit of the transmitted frequency,
t_s varies between 0 and T_s,
T_s is the repetition period of the modulation function.

At any instant, $t = t_n + t_s$, where t_n is the beginning of the nth period (see Fig. 3).

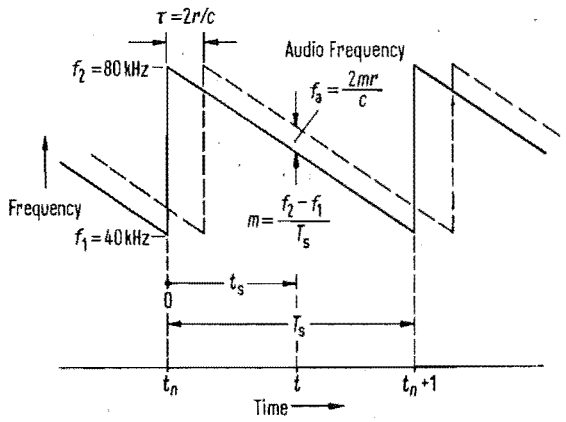


Fig. 3. Transmission and reflected signal parameters.

The interaural amplitude difference IAD(t) is also a function of time under conditions of motion and is proportional to $\theta(t)$. Thus when these spatial codes are used as auditory stimuli the resolution capability must be a function of four variables

$$[f_{a_1}(t), \mathrm{IAD}_1(t), f_{a_2}(t), \mathrm{IAD}_2(t)].$$

Clearly any function of four dynamic variables is too complicated to be studied and described by subjects at this stage. Hence in the experiment to be described f_{a_2} IAD_1 and IAD_2 are held constant, varying only f_{a_1} as described in eq. (1). The experimental procedure was based entirely upon the "psychological phenomenon" described below.

3.1. The psychological meaning of auditory resolution under conditions of change

One of the difficulties in psychophysical experiments is to determine a suitable method of verifying the psychological judgements of subjects to ensure that they respond to the psychophysical phenomena to be studied. It was known [5] that presenting simultaneously two tones assigned values f_{a_1} and f_{a_2} each with its own value of IAD produces a result which is similar to that obtained in the experiment described in Section 2. The assignment of a value of IAD (by Rowell) did not affect the capability of resolution in range under static conditions as compared with zero IAD as used in the experiment of Section 2, but it did more clearly define the psychophysical phenomena being studied compared with Rowell's earlier experiments, since subjects were required to indicate *left* or *right* directions to the second tone. The two tones had to be perceived separately to do this.

This same decision was required in the dynamic experiment. It was designed to simulate conditions experienced when using the sonars. Initially at $t = 0$, $f_{a_1}(t) \cong f_{a_2}$ and the subject hears slow beats and feels only one complex image in the auditory space. Its IAD is the combination of IAD_1 and IAD_2. As t increases $f_{a_1}(t)$ decreases according to eq. (1). The interference between the two tones changes in a unique way (because of the signal uniqueness) until at a certain instant $t = t_1$ the subject feels the image of $f_{a_1}(t)$ flow "left" or "right" in auditory space, according to the assigned values of IAD_1 and IAD_2. The attention appears to be captured by the "moving" tone which may be thought of as forming a flow pattern in space. At this instant the subject is required to respond by pressing a switch to store the value of $f_{a_1}(t)$. $f_{a_1}(t)$ is allowed to continue when the attention can be transferred to f_{a_2} which then appears in its appropriate auditory spatial position. This is always secondary to the initially perceived flow of $f_{a_1}(t)$, but occurs almost simultaneously. The psychological phenomenon is described in Fig. 4. The judgment of subjects can be verified by questioning them on the relative movement of the images. The frequency resolution can be defined here as the frequency difference between f_{a_2} and $f_{a_1}(t)$ at which the two tones separate and "slide" in auditory space to their respective positions.

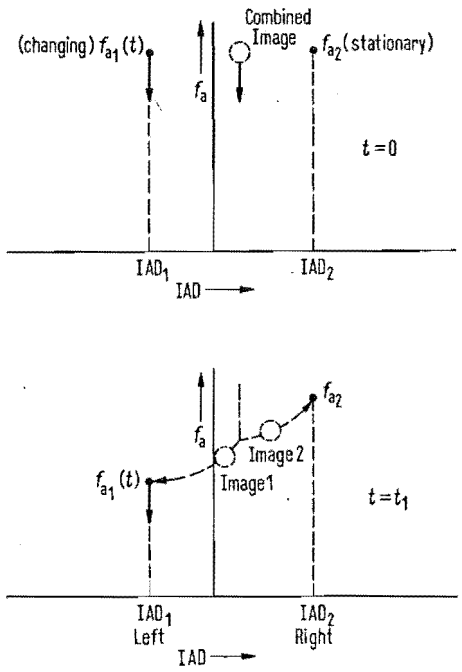


Fig. 4. The psychological phenomenon of resolution in a changing situation.

3.2. Simulation of the problem

We are concerned here with simulating a situation which relates to a real experience. In the experimental situation described, the auditory sensations resemble what would be obtained if one object at range $r_1\theta_1$ approached the sonar with constant radial velocity v, i.e. $r_1(t) = r_1(0) - vt$ and a second object were stationary at a range

$$r_2\,[= r_1(0),]\,\theta_2.$$

Then the range coding frequency is given by

$$f_{a_1}(t) = \frac{2m r_1(0)}{c} + \frac{2v}{c} f_2 - \frac{2v}{c} m t_s - \frac{2v}{c} m t \qquad (2)$$

$$0 \leqq t_s \leqq T_s$$

or

$$f_{a_1}(t) = f_{a_1}(0) - \frac{2v}{c} m (t + t_s). \qquad (3)$$

A linear voltage controlled oscillator was used to produce the frequency function from the voltage function

$$V_{a_1}(t) = V_{a_1}(0) - \alpha(t + t_s) \quad (4)$$

where α and $V_{a_1}(0)$ are linearly related to $2vm/c$ and $f_{a_1}(0)$. An analogue computer was used to produce the voltage function of eq. (4) and to control the experiment.

Fig. 5 shows the panel patching of the computer and Fig. 6 shows the arrangement of the equipment for the experiment.

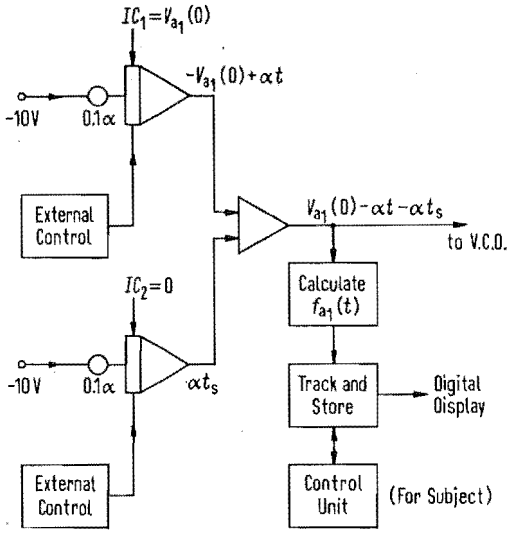


Fig. 5. Analogue computer to generate $V_{a_1}(t)$ and to control Experiment 2.

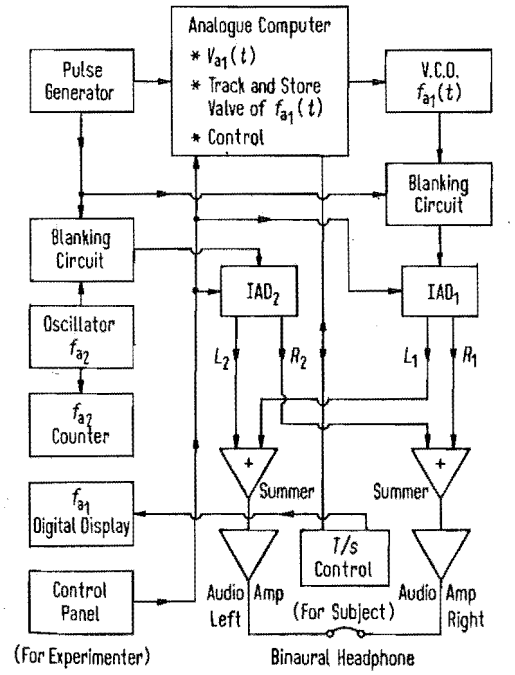


Fig. 6. Arrangement of equipment for Experiment 2.

3.3. Experimental procedure

The subject's task was to listen to the complex sound including $f_{a_1}(t)$ and f_{a_2} as described and press the switch to store the value of $f_{a_1}(t)$ at the instant the descending tone was perceived to slide to the left or the right of the initial complex image. Each subject was questioned about the direction of shift of either the low tone ($f_{a_1}(t)$) or the high tone (f_{a_2}). For each value of f_{a_2} the experimenter recorded the values of $f_{a_1}(t)$ for different values of IAD_1 and IAD_2. These latter are changed in a random manner according to a table.

3.3.1. Experiment 2 (The simulation relates to the fish sonar because the rates of change encountered are such as to allow a subject to respond without introducing serious error due to the delay in response)

Suppose that the approach velocity of object 1 is $v = 1.5$ m/s the velocity of sound is taken as 1500 m/s $T_s = 3.2$ s, and $m = 12.5$ kHz/s then

$$f_{a_1}(t) = f_{a_1}(0) - 25\,(t + t_s) \quad 0 \leqq t_s \leqq 3.2\,\text{s}. \quad (5)$$

Three subjects S_1, S_2, S_3, were used in the experiment. Values of $f_{a_1}(0) = f_2$ were chosen to be 1000, 2000, and 3000 Hz, while the values of IAD were 0, +4, +8, +12 dB for object 1 and −10, −6, −2, +2, +6, and +10 dB were chosen for object 2.

In this experiment a delay in stopping $f_{a_1}(t)$ of say an upper limit of 0.5 s, due to a subject's reaction time, may introduce an error of +25 Hz to the recorded value of $f_{a_1}(t)$. However, since $f_{a_1}(t)$ decreases in a saw-tooth manner with a sweep of 160 Hz per sweep period T_s and a reset of 80 Hz (see Fig. 7), the subject's delay may occur at the end of

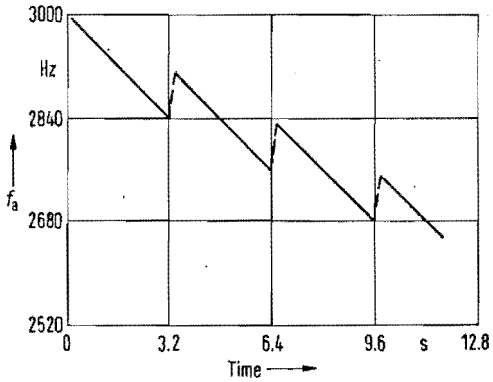


Fig. 7. Variation of range coding frequency with respect to time in Experiment 2.
$f_{a_1}(t) = 3000 - 25t - 25t_s$, where $0 \leqq t_s \leqq 3.2$. The broken lines correspond to ambiguous signals blanked out.

a sweep period when $f_{a_1}(t)$ is reset 80 Hz to start a new sweep. The error may then be (25 — 80) Hz = — 55 Hz. To reduce these errors $f_{a_1}(t)$ was recorded four times and averaged. Fig. 8 shows the results of the three subjects. The variation of $f_{a_1}(t)$ due to delay and reset errors and subject variation was greater than the variation between values of IAD. The dependence of $f_{a_1}(t)$ on the IAD cannot therefore be determined. It is now assumed to be negligible so the value of Δf for each f_{a_2} is found by averaging over all values IAD_1 and IAD_2.

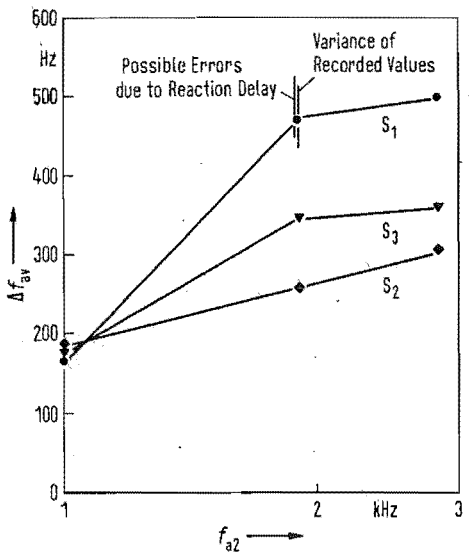


Fig. 8. Variation of the frequency resolution Δf_{av} with respect to frequency averaged over all values of IAD (IAD variable has negligible effect).

Quite a wide variation between subjects can be expected, but "within" subject variation is seen to be small.

3.3.2. Experiment 3 (Frequency resolution with cyclic variation removed)

We wished to determine the influence of the saw-tooth variation on the resolution capability of the auditory system. Whilst in practice this cannot be eliminated, and it plays an important part in sharpening the ambiguity function of the sonar system, we could not predict its influence on the auditory resolution. The frequency function for this experiment became

$$f_{a_1}(t) = \frac{2m}{c} r(t) = f_{a_1}(0) - \frac{2v}{c} m t \qquad (6)$$

from which we get

$$f_{a_1}(t) = f_{a_1}(0) - 25\,t. \qquad (7)$$

Repeating the experiment with S_2, who demonstrated a higher resolution capability in Experiment 2, the following result was obtained.

Table I.
Frequency resolution with cyclic variation removed.

f_{a_2} Hz	Δf Hz	Increase in Δf %
1000	220	170
2000	440	170
3000	630	210

3.3.3. Experiment 4 (The effect of increasing the approach velocity

Since the saw-tooth had a significant effect on the resolution capability, we wondered what the effect would be of increasing the rate of change of frequency. This obviously increased the possible error due to delay in subject's response. The object velocity was doubled to 3 m/s doubling the slope of $f_{a_1}(t)$ and the reset value as seen in Fig. 6. The following result was obtained again by S_2.

Table II.
Effect of increasing velocity on frequency resolution.

f_{a_2} Hz	Δf Hz	Decrease in Δf %
1000	130	30
2000	230	11
3000	240	16

Table III.
Comparison of frequency resolution in four experiments.

f_{a_2} Hz	$\Delta f(1)$ Hz	$\Delta f(1)/f$	$\Delta f(2)$ Hz	$\Delta f(2)/f$	$\Delta f(3)$ Hz	$\Delta f(3)/f$	$\Delta f(4)$ Hz	$\Delta f(4)/f$
1000	400	0.50	180	0.20	220	0.25	130	0.14
2000	700	0.42	250	0.13	240	0.25	230	0.12
3000	1100	0.45	300	0.11	630	0.23	240	0.08

$\Delta f(1)$ simultaneous presentation of f_{a_1} and f_{a_2}
$\Delta f(2)$ saw-tooth frequency decrease $f_{a_1}(t)$
$\Delta f(3)$ linear frequency decrease $f_{a_1}(t)$
$\Delta f(4)$ saw tooth frequency decrease $f_{a_1}(t)$ at double rate.

Quantitatively the improvement in resolution is not significant except at low frequency. Qualitatively however, the sensation of image slide is more pronounced and the separation of images is very clear by all who experience the phenomenon. S_2 was unable to respond to the sensation sufficiently quickly to change greatly the measured resolution capability obtained in Experiment 2.

Summarising resolution of S_2:

4. Summary of results

From the above experiments, results can be compared and summarised as follows:

1) The capability of frequency resolution, under dynamic conditions is significantly improved when compared with that under stationary conditions.

2) The frequency resolution, measured in the dynamic case, appears to be independent of the values of IAD_1 and IAD_2. Rowell [5], when determinining the frequency difference between two stationary tones so that each could be localised, also reported this non-relationship; however, he also showed a frequency difference of almost an octave.

3) In CTFM sonar, the Doppler effect results in an increase of the rate of change of the range coding frequency by a factor of two, within each sweep cycle. This gives an improvement in the resolution capability (compare experiments 2 and 3).

4) An increase in the relative velocity of target and sonar also improves the resolution capability of the auditory display system, quantitatively as well as qualitatively (comparing examples 2 and 4). However, if the target velocity results in too large a rate of change of the range coding frequency, the human reaction time becomes an important factor in the measurement of resolving power and the tracking of objects. Thus, in the Sonar for the Blind when $T_s = 0.250$ s, $v = 1.5$ m/s and $r_1(0) = 3$ m, signals will be resolved cognitively long before reaction can be recorded.

(Received August 5th, 1975.)

References

[1] Kay, L. and Do, M. A., An artificially generated multiple object auditory space for use where vision is impaired. Acustica **36** [1976], 1.
[2] Kay, L., Sonic glasses for the blind — a progress report. A. F. B. Res. Bull. **25** [1972], 25.
[3] Kay, L., Sonic glasses for the blind — presentation of evaluation data. A. F. B. Res. Bull. **26** [1973], 35.
[4] Kay, L., Towards objective mobility evaluation — some thoughts on a theory. A. F. B. monograph, New York 1974.
[5] Rowell, D., Auditory display of spatial information. Ph. D. Thesis, University of Canterbury, New Zealand 1970.
[6] Smith, R. P., Transduction and audible displays for broad band sonar systems. Ph. D. Thesis, University of Canterbury, New Zealand 1973.
[7] Plomp, R. and Steeneken, H. T. M., Interference between two simple tones. J. Acoust. Soc. Amer. **43** [1968], 883.